DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS FOR STOCHASTIC SYSTEMS

YIN JUN
(*B.Eng., University of Science and Technology of China*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

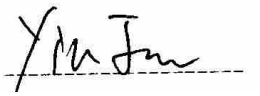
DEPARTMENT OF INDUSTRIAL & SYSTEMS ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2012

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



YIN JUN

4 June 2012

Acknowledgements

First and foremost I offer my sincerest gratitude to my supervisor, A/Prof. NG Szu Hui, who has supported me thoughout my Ph.D study with her patience and encourage. I'm grateful for her suggestions and comments to all of my research work. All these would not have been possible without her efforts.

I also would like to thank my co-supervisor, A/Prof. NG Kien Ming, for his kindly guidance and valuable suggestion during the writing of this thesis.

My parents support me thoughout my entire study in China and Singapore. I was away from home for a long time and could not take good care of my family. I would like to offer my sincerely gratitude and love to them.

All of my classmates and friends in Singapore, CHEN Ruifeng, HAN Dongling, LV Yang, LIU Xiangjun, LIU Jin, MU Aoran, XIONG Chengjie, YU Jinfeng, SHENG Xiaoming and Dr. Lim Yee Nah from NUH , I couldn't get through without your encourage and help.

Last but not the least, I want to say thank you to my wife ZHONG Ying, for all her understanding and support to my work.

Abstract

This thesis studies the design and analysis of computer experiment for stochastic simulations. The stochastic simulation models play an important role in modern industrial and managerial applications. However, its stochastic response increases the difficulties of conducting analysis and experiments. This thesis proposes the kriging metamodel with modified nugget effect as a solution to the more general stochastic simulation scenario with hetergeneous variances. The results suggest that the proposed model performs beter than the existing models by appropriately account for the influence of random noise in terms of model prediction and parameter estimation. The study on parameter estimation uncertainty problem with kriging metamodels in stochastic simulation is further investigated. Based on the proposed model, a two-stage optimization algorithm is also developed as the solution to stochastic simulation optimization for heteroscedastic case. The numerical results suggest that the proposed model can effective reduce the erratic behavior of the predictor by more appropriately accounting for the influence of the stochastic responses. Last, a Bayesian metamodeling and two-stage sequential design approach are also developed to overcome the parameter estimation uncertainty issue and efficiently use the limited computing budget in practice.

Keywords: simulation, metamodels, optimization, design of experiment, stochastic systems, discrete event simulation

Contents

1 INTRODUCTION **1**

1.1 Computer Simulation Model and Computer Experiments 1

1.2 Deterministic Simulation Model and Computer Experiments . . . 3

1.3 Stochastic Simulation Model and Computer Experiments 5

1.4 Objective and Scope 8

1.5 Organization 9

2 LITERATURE REVIEW **12**

2.1 Review of Metamodels 12

2.1.1 Polynomial Regression Model 12

2.1.2 Spatial Correlation Model 13

2.1.3 Multivariate Adaptive Regression Splines Model 14

2.1.4 Radial Basis Function Model 15

2.1.5 Artificial Neural Network Model 16

2.2 Review of Kriging Metamodel in Computer Experiments 17

2.2.1 Kriging Metamodel in Homoscedastic case 18

2.2.2 Kriging Model in Heteroscedastic case 20

2.3 Review of Design of Experiment for Computer Simulation 22

2.3.1 Space-filling Designs 23

2.3.1.1 Latin hypercube design 23

2.3.1.2 Uniform design 24

2.3.1.3 Distance dependent design 25

2.4 Designs Based on Optimization Criterion 25

2.4.1 Response surface methodology 25

2.4.2 Trust region method 26

2.4.3 Efficient global optimization 27

3 KRIGING METAMODEL WITH MODIFIED NUGGET EFFECT 29

3.1 Introduction 29
3.1.1 Differences from the stochastic kriging model 34
3.1.2 Organization 35
3.2 Kriging Model with Modified Nugget Effect 35
3.2.1 Classic kriging (deterministic and nugget effect model) . . 35
3.2.2 The development of kriging metamodel with modified nugget effect 38
3.2.3 Parameter estimation and characteristics of likelihood function with noisy data 42
3.2.4 Error measurement 47
3.3 Prediction Performance of the Kriging Model with Modified Nugget-effect 48
3.3.1 Comparison through MSE_S 48
3.3.2 Estimating predictor's variance 49
3.4 Examples 52
3.4.1 Test Function 52
3.4.2 M/M/1 queueing system 57
3.4.3 PAD system 59

4 PARAMETER ESTIMATION FOR KRIGING METAMODEL IN STOCHASTIC SIMULATION 67

4.1 Introduction 67
4.2 Decomposition of the Overall Prediction Error for Stochastic Case 69
4.3 Maximum Likelihood Estimation with Stochastic Response 71
4.3.1 A simple two-point problem 72
4.3.2 Analytical Results 73
4.3.3 Influence of Parameter Estimation on Overall Prediction Error 75
4.4 Numerical Experiments 76
4.4.1 One Dimension Quadratic Test Function 76

4.4.2 Two Dimension Linear Function 78
4.4.3 Two Dimension Sinusoidal Function 79

5 OPTIMIZATION OF STOCHASTIC SIMULATIONS WITH KRIGING METAMODEL 84

5.1 Introduction 84
5.2 The expected improvement function 86
5.3 Limitations of EGO and SKO in Noisy Heteroscedastic Situations 87
5.3.1 Characteristics of Good Algorithms and Criteria 91
5.4 Development of Methodology 92
5.4.1 The search stage 93
5.4.2 The allocation stage 93
5.4.3 An algorithm overview 95
5.5 Numerical Examples 100
5.5.1 Single dimension test function(Comparative study) 100
5.5.2 Two Dimension Keys and Reese (2004) Function (Comparative Study) 102
5.6 Ocean Liner Example 106
5.7 Conclusion 111

6 BAYESIAN METAMODELING AND DESIGN APPROACH FOR STOCHASTIC SIMULATIONS 115

6.1 Introduction 115
6.2 Model Formulation 118
6.2.1 Modeling Uncertainty 119
6.2.2 Observed Data 120
6.2.3 Bayesian Prediction and Predictive Distribution 120
6.2.3.1 Derivation of the Predictive Distribution (*Assuming ϕ_Z is known*) 121
6.2.3.2 Modeling of σ^2_ξ 121
6.2.3.3 A further simplification of Equation (6.6) 124
6.2.3.4 A General Approach to Deriving the Predictive Distribution (when all parameters are unknown) . 127
6.3 Numerical Examples 128

6.3.1 The Simple Quadratic Function 128
6.3.2 The M/M/1 System 132
6.4 Sequential Experimental Design Approach 135
6.4.1 The two stage design framework 135
6.4.2 A follow-up design criterion 136
6.4.3 Simplification and decomposition of the IMSPE 137
6.4.3.1 A simplified Stage 2 design for the two point example 139
6.4.3.2 A numerical study on the EIMSPE for different design options 142
6.4.4 Improved two-stage design approaches 146
6.4.4.1 One-Point-at-A-Time (OPAT) sequential design approach 146
6.4.4.2 Simple two-stage design approach 148
6.5 Comments and Conclusions 150

7 CONCLUSION **152**
7.1 Main findings 152
7.2 Future research 154

References **169**

A Kriging predictor and kriging variance for heteroscedastic model **170**

B MSE for the modified nugget-effect and nugget-effect model **172**

C Proof for the two-stage algorithm **175**

D Details of the two-point example **178**

E Estimating Predictor Variance by Delta Method **180**

F Proof for Proposition 1 **182**

G Posterior distribution of the parameters **184**

H Posterior distribution of σ_Z^2 188

List of Figures

3.1 Test function with step variance function. 31
3.2 Ordinary kriging and nugget-effect model for the test function. . . 32
3.3 Likelihood function for ϕ_Z (signal function only and noisy observation in Equation (3.1)). 43
3.4 Likelihood function for ϕ_Z with nugget effect model (noisy observation of the signal function). 44
3.5 Profile of the penalized portion of the likelihood function for modified nugget effect model. 46
3.6 Different predictors' output for test function (r_{var} =10). 53
3.7 Different predictors' output for test function (r_{var} =100). 54
3.8 Different predictors' output for test function (r_{var} =1000). 55
3.9 Different predictors' output for test function (r_{var}=100, 2nd-order polynomial regression model). 56
3.10 Influence of nugget value on MSE (test function). 57
3.11 Studentization method with 100 sub-groups (sample size per sub-group = 10. 59
3.12 Modified nugget effect model with 100 sub-groups (sample size per sub-group = 10). 60
3.13 Queueing model for computer PAD system. 61
3.14 Prediction interval for nugget effect predictor (PAD system). . . . 62
3.15 Prediction interval for modified nugget effect predictor (PAD system). 63

4.1 Design for two-point problem. 72
4.2 Two dimension linear test function. 79

4.3 Two dimension sinusoidal test function. 81
4.4 Design of the sinusoidal test function. 82

5.1 EI function and response metamodel with noisy test function (white noise) 88
5.2 AEI function and response metamodel with noisy test function (white noise) 89
5.3 Modified AEI function and response metamodel with noisy test function (non constant variance) 90
5.4 Contour plot of EI function of predictor mean difference and standard deviation using the Modified Nugget Effect Kriging model . 91
5.5 $r_A(i)$ at different iterations as $\frac{I}{n_0}$ changes. 99
5.6 One dimensional example with proposed algorithm. 102
5.7 One dimensional example with proposed algorithm.(Higher variance) 113
5.8 Contour plot of the two dimension test function. 114
5.9 AEX service route(distances in nautical miles). 114

6.1 Average plug-in MSPE of MNEK and the observed MSPE for the high variance scenario. 118
6.2 Predictive mean given by MNEK, BKS and BKMCMC for the simple quadratic function example for the low and high variance scenarios (with (a) constant and (b) quadratic mean functions). . 129
6.3 Predictive variance given by MNEK, BKS and BKMCMC for the simple quadratic function example (with (a) constant and (b) quadratic mean functions). 130
6.4 The influence of the random noise level on the optimal location of the new design point for the two point case with $\phi_z = 0.5$. 142
6.5 Υ with $\phi_z = 0.01, 0.5, 1$. 143
6.6 Results for Option 1 (solid line) and Option 2 (dotted line) for the eight test scenarios. 144

Chapter 1

INTRODUCTION

This thesis contributes to the design and metamodeling methods for the Design and Analysis of Computer Experiments(DACE) for stochastic systems. In this chapter, we first briefly introduce the background and development of the computer simulation model and computer experiments in Section 1.1. Following, Section 1.2 will trace the development of metamodels, DACE for deterministic systems. Section 1.3 will review the development and current progress on the research of DACE for stochastic systems, and the gaps of the current research will also be highlighted in this section. Based on the gaps specified in Section 1.3, the objective and scope of this thesis will be provided in Section 1.4.

1.1 Computer Simulation Model and Computer Experiments

A computer simulation model is a computer program that attempts to simulate the behavior of a specific actual system. The use of computer simulation model provides a effective and efficient way to study and analyze complex systems which have no closed form solution and require intensive computational effort. Example of computer simulation model can be found in a variety of science and engineering field. Early applications could date back to the Manhattan Project in World War II. Currin *et al.* (1991) presented a integrated circuit simulation model and the related design of experiment issues. Computer simulation model is also ap-

plied in meteorological and environmental research, see Watson & Johnson (2004) and Chin & Melone (1999). Computer simulation softwares based on the Finite Element Method(FEM) are popular in Computer Aided Design(CAD) for many engineering design problem, such as COMSOL Multiphysics, CST, HFSS etc. Rao & Balakrishnan (1999) gave a inclusive review on computational techniques and computer simulation's applications in electromagnetic engineering design problem. Computer simulation models are also well applied in assessing changes in operations and managerial policies, see Greenwood *et al.* (2005) and Yao *et al.* (2011).

The needs of computer simulation models naturally leads to the study of computer experiments, which refer to the experiments conducted on computer simulation models. Similar to the experiments conducted on the real world physical systems, computer experiments refer to changing the inputs of system and observing the corresponding system outputs. With these input/output data combinations, the researcher can study the inner mechanism or behavior of the target system, which is very helpful for the analysis of complex systems with no analytical closed form solutions. Compared with the physical experiments, conducting experiments on computer simulation models has several benefits:

1. Computer simulation models usually are comparatively cheaper and easier to build and execute.

2. Computer experiments are based on the computer program, hence it mainly limited by the computational capability.

However, the computer models and computer experiments also have some limitations, such as whether the computer simulation model can imitate the actual physical system with a satisfactory accuracy level. Hence, the validation and calibration of the computer simulation models are essential for the actual practice. How to reduce the differences between the finding of computer experiments and the true mechanism of the real world systems becomes the key problem for the research of computer experiments.

Computer simulation models can be categorized in different ways, such as steady-state or dynamic, continuous or discrete, etc. One widely accepted categorization method is to divide the computer simulation models as deterministic or

stochastic simulation models. Unlike the real physical systems, the deterministic simulation model always generates the exactly same outputs given the fixed inputs. However, the stochastic simulation model contains randomness just as the real physical systems. This difference between the deterministic and stochastic simulation models leads to different design and analysis approaches for computer experiments. In the next section, we will first look into the development of deterministic simulation model and computer experiments.

1.2 Deterministic Simulation Model and Computer Experiments

Deterministic simulation model are commonly used in the cases where underlying mechanism or averaged behavior of the target system is of our interest. In these cases, the randomness of the real physical system usually has low impact on the system's performance. Examples can be found in Computer Aided Engineering (CAE) and Computer Aided Design (CAD), see Kleijnen (2008) and Santner *et al.* (2003). Deterministic simulation models become a popular approach for many modern engineering design and product development problems due to its convenience and comparatively lower cost. However, as the complexity of the simulation model increases, the computational cost of running the simulation model become the critical issue. To simplify the problem and reduce the cost, one common practice is to build a simplified metamodel, or surrogate model for the simulation model. Metamodel is a closed form mathematical model that can imitate the behavior of simualtion model with less computational effort. For the choice of metamodels, the most common technique has been based on the parametric polynomial response surface approximations. Although polynomial response metamodels offer good approximations for simple cases, the main drawback of the polynomial metamodels is their lack of flexibility to achieve a global fit for complex cases. To account for the high nonlinear responses of complex simulation models, various metamodels like the kriging, multivariate adaptive regression splines (MARS), radial basis function (RBF), artificial neural networks (ANN), and support vector regression (SVR) have been proposed in recent years.

Reviews of these metamodels' performance and applications in engineering can be found in Simpson *et al.* (2001) and Li *et al.* (2010a).

Among all types of these metamodels, the kriging metamodel is one of the more promising metamodels. The kriging metamodel is originated from the mining technology and geo-statistic, see Matheron (1963). It was introduced into the computer experiment by Sacks *et al.* (1989) and quickly became a popular model in the field. The kriging metamodel has been successfully applied to many deterministic computer experiments as its interpolating characteristic is appropriate for the deterministic case. It is more adaptable than the regression based models and not as complicated and time consuming as artificial intelligence techniques.

For the design of computer simulation with deterministic outputs, as mentioned in Section 1.1, the experimental design for the deterministic simulation model is different from the DOE for the real physical systems. For example, Santner *et al.* (2003) mentioned that the commonly used techniques in physical experiments like randomization, blocking and replication methods are usually not adopted for a typical deterministic simulation experiment since its output always stay the same given the same input. According to Santner *et al.* (2003), one of the most important type of design method for deterministic computer experiment is the space-filling designs. Space-filling designs have several benefits for the application in deterministic computer simulations: First, each of the design point for the space-filling design is unique, which is reasonable as replication would not provide additional information for deterministic computer experiment. Second, space-filling design assumes that every parts of the design space have the equal importance, which helps in spreading the design points evenly out in the whole design space. For the space-filling design, the Latin Hyper Cube Design (LHD), Min-max and Max-min design, uniform design are commonly used. With the random sampling techniques, distance criterion or the uniformly located design points, all these design approaches intends to spread out the locations of the design point in the entire sample space, hence the metamodel can be capable of universally capturing the behavior of the computer model. For the deterministic computer simulation, the key is the location of the input x as the computer model itself is deterministic, which means the locations of the inputs will determine the output of the computer model.

In applying kriging metamodel as a surrogate for optimizing the deterministic simulation model, a sequential approach is typically taken. Jones *et al.* (1998) proposed a sequential optimization method based on the Kriging metamodel and the Bayesian Global Optimization approach. The proposed method applied the Expected Improvement (EI) function and the Efficient Global Optimization (EGO) algorithm to balance the local and global search for the optimum of an unknown response surface as the solution to the global optimization of the corresponding deterministic simulation model. This method is a Kriging metamodel based optimization method developed from the Bayesian based optimization methods in Mockus (1994). Kleijnen & Beers (2010) extends this sequential optimization approach by introducing an improved estimator of the kriging variance through bootstrapping. As the originally proposed EI function and EGO algorithm are designed for deterministic scenarios, it considered the allocation of the design points as the only design option for experimenter and focused on balancing the search within the local area of the current optimum and the entire sample space. However for stochastic simulations, the random variability of the stochastic response can considerably affect the metamodel fit (Yin et al. 2009) and therefore the search for the optimum. In this situation, the experimental design is further affected by the stochastic noise in the simulation. Hence, in addition to reducing the spatial uncertainty by observing new design points, the experimenter must also consider the influence of random noise.

1.3 Stochastic Simulation Model and Computer Experiments

Unlike the deterministic simulation models, stochastic simulation models assume randomness in the outputs. Researcher usually use the stochastic simulation model to represent the real world randomness, such as uncontrollable factors in chemical reactions, weather phenomenon or market fluctuation. Examples can be found in fields like operation research, economic study or financial engineering, see Asmussen & Glynn (2007). Compared with deterministic simulation model,

the stochastic simulation model is closer to the realistic, and hence more suitable for short term forecasting, social behavior related applications and etc.

Due to the randomness and complexity of the stochastic simulation model, the cost of conducting experiment on the simulation model can be very expensive. Hence the metamodels and experimental design techniques are popular for these years. To be specified, stochastic computer experiments can be divided into two different scenarios: homoscedastic case and heteroscedastic case. The homoscedastic case refers to the situation where the random noise in the stochastic computer simulation is assumed to be Normally, Independently and Identically distributed (NIID), which can be appropriately modeled by some existing kriging models, like the kriging model with nugget effect, see Cressie (1993) and Huang *et al.* (2006). These models and methods are very successful when the underlying homoscedastic assumptions are met. However, the performance deteriorates fast when the noise varies, see Yin *et al.* (2008) and Li *et al.* (2010a). Existing research like Kleijnen & Beers (2005) proposed methods to transfer the heteroscedastic case into the homoscedastic case or even deterministic case where the traditional kriging metamodel is applicable. These methods however need sufficient computing budget and prior information about the random noise. For the more general stochastic computer experiments with heterogeneous variance, a suitable model has yet to be found.

For the computer experiments with the stochastic simulation model, the basic idea is close to conducting experiment on the real physical systems due to the existence of randomness. Techniques like replication, blocking and randomization can be used. There are some existing experimental design approaches and optimization methods for stochastic simulation, including the sequential Response Surface Methodology (RSM), see Angün *et al.* (2002), the Stochastic Approximation (SA) method, see Kushner & Clark (1978), the Nested Partitions (NP) method, see Shi & Olafsson (2000), and other heuristic methods like the Genetic Algorithm and Simulated Annealing. Tekin & Sabuncuoglu (2004) provides a comprehensive review of the different approaches for simulation optimization.

Huang *et al.* (2006) adapted the EGO scheme for stochastic simulation models and proposed the Sequential Kriging Optimization (SKO) method for optimizing stochastic systems. With the nugget effect Kriging model and augmented EI

function, the SKO algorithm accounts for the influence of random noise. However, SKO only considered the homoscedastic cases where the random noise function are assumed to have constant variances throughout the entire sample space. For the more general case with heterogeneous variances, SKO is unable to capture the behavior of the stochastic simulation model due to the mis-specified assumption on the variances of the stochastic response. Hence the estimated global optimum obtained by the SKO with augmented EI function can be far away from the true optimum due to the inadequate fit of the Kriging model. Picheny et al. (2010) extended the EI based optimization algorithm to the case with normally distributed noise and non constant variances. In addition, they proposed a more general quantile-based criterion, Expected Quantile Improvement (EQI) to take into account the user's risk tolerance. The higher the user sets the quantile, the more conservative the criterion will be and vice versa. Their algorithm accounts for limited computing budget and also considers the variance of the noise at unsampled locations when searching for a new point. This gives the algorithm a desirable characteristic of favoring exploration at the start where available budget is high, and becoming more conservative towards the end. However, it also requires the noise variance function be known, and the algorithm's computational complexity is greater compared to traditional EI. In addition, Picheny et al. (2010)'s algorithm with the online allocation does not allow backtracking, meaning once a point has been selected by the criterion and sampled until a condition is met, that point is never re-visited again. In an iterative algorithm where more and more information about the objective function is revealed as the algorithm progresses, this characteristic may not be ideal.

Clearly, the metamodel designed for the deterministic simulation needs to be improved in order to take accounts of the stochastic response. Making homoscedastic assumptions on the random noise component for the model, the kriging metamodel can be developed into kriging metamodel with nugget effect (the nugget effect model), see Cressie (1993). However, the appropriate model is still missing for the more general heteroscedastic case. Existing methods including the replication method and studentization method proposed by Kleijnen & Beers (2005) essentially converts the general heteroscedastic problem into the homoscedastic problem, then the deterministic kriging model or the nugget effect

model can be applied to the problem. However, these type of methods require prior information of the simulation model and sufficient computing budget to reduce the variability of the observed data, which is unrealistic for most of the real-world cases. As a result, it naturally leads to the issue of developing a suitable model for the heteroscedastic case.

1.4 Objective and Scope

As indicated in the previous section, the gaps for current research in the field of computer simulation for stochastic system can be summarized as follows:

- The existing kriging model with nugget effect is designed for the homoscedastic case. In order to apply the nugget effect model in the more general heteroscedastic case, the heteroscedastic case has to be transformed into the homoscedastic case. This transformation usually needs considerable additional computing budgets. However, the computing budgets are always seriously limited for most of the real world problems.

- There are limited studies of the parameter estimation stochastic simulation so far. More specifically, in the stochastic simulation environment, the parameter estimation uncertainty of the model estimation is not appropriately accounted for.

- For the experimental design issue, experimental design for stochastic simulation with heterogeneous variance has additional allocation problems compared with the experimental design for stochastic simulation with homogeneous variance. This has to be considered in the more general experimental design method for stochastic simulation.

This thesis intends to present a novel kriging model and experimental design approach adapted for the general stochastic simulation with heterogeneous variance. The objectives of this research are to:

- Extend the existing kriging model to the modified kriging model in order to appropriately account for the random noise with heterogeneous variance.

- Investigate the effect of random inputs with high variability on the parameter estimation uncertainty for kriging model and compare the performance on parameter estimation for different kriging models.
- Develop the experimental design for the more general stochastic simulation with heterogeneous variance. Both of the sensitivity analysis and optimization criterion in the design should be considered.

The result of this study may provide an alternative solution for DACE in stochastic simulation, especially for the heteroscedastic case. Moreover, this study may help in increasing

- The understanding of the stochastic simulation model and kriging model's behavior.
- The robustness of the parameter estimation for kriging model in stochastic simulation circumstance.
- The performance and efficiency of the experimental design for stochastic simulation.

One shortcoming of the kriging model is that it cannot handle high dimension inputs, as the high dimension data will significantly increase the scale of the correlation matrix inside the model and difficulty of resolving the equations. However, since this research mainly focuses on the behaviors of the stochastic simulation, the data dimension is not central to this study. As a result, we only focus on the low dimension data in this study.

1.5 Organization

This thesis contains 7 chapters. In Chapter 2, literatures related to this research will be reviewed. The review is going to be separately provided for both the metamodels, designs of experiment and metamodel based optimization method. For the metamodel part, we focus on the more promising kriging metamodel which is the model proposed to applied in the following studies.

In Chapter 3, the kriging model with modified nugget effect is proposed as the solution to the general stochastic simulation with heterogeneous variance. We develop the model on the basis of the kriging model with nugget effect by relaxing the homoscedastic assumption on the noise process, and we provide the comparison among the predictors' forms among differen kriging models. Moreover, we further investigate the differences between the proposed model's performance and the deterministic kriging model by analyzing the influence of the random noise on the parameter estimation uncertainty of the model. Other than the kriging predictor, we also study the estimation of the variance of noise process at unobserved location with different methods. Finally, numerical examples are presented to illustrate the differences between the propose kriging model with modified nugget effect and existing methods.

In Chapter 4, we further extend the research on the parameter estimation uncertain for kriging model with heteroscedastic noise in the Chapter 3. The overall prediction error of the kriging predictor is decomposed into three parts: model misspecified error, prediction errors caused by random noise and parameter estimation uncertainty. We use a simple two-point example to theoretically illustrate the random noise's influence on the parameter estimation and further explain in detail that the kriging model with modified nugget effect can compensate this parameter estimation uncertainty. Three numerical test functions are also provided as the examples indicating the differences between different kriging models in terms of the decomposed prediction errors.

In Chapter 5, we apply the proposed kriging model with modified nugget effect to the design of experiment for the stochastic simulation with heterogeneous variance. Based on other kriging model based method like the Efficient Global Optimization (EGO) and Sequential Kriging Optimization (SKO), we propose the two-stage sequential design framework together with the modified nugget effect kriging model as the alternative method for the heteroscedastic case. The two-stage framework is design to better balance the different design options that the experimenter might face in the stochastic scenario with non constant variance. We also accordingly modify the Expected Improvement (EI) function to better account for the influence of the random noise with non constant variance. The EI function is adopted in the previous studies to evaluate the potential value

of the unobserved points in terms of the design locations. We proposed several different types modified EI functions to account for both the influences of unobserved points and random variability in different stochastic scenarios. Simple test examples are used to show the way that new two-stage sequential framework performs. A more realistic shipping liner planning simulation model is also adopted as an example to demonstrate the usage of the proposed design framework and modified nugget effect kriging model in the real world practice.

In Chapter 6, we propose a Bayesian metamodeling approach for kriging prediction is for stochastic simulations to more appropriately account for the parameter estimation uncertainties mentioned in Chapter 4. We derive the predictive distribution under certain assumptions and also provide a general Markov Chain Monte Carlo analysis approach to handle more general assumptions on the parameters and design. Numerical results indicate that the Bayesian approach has better coverage and closer predictive variance to the empirical value than a previously proposed modified nugget effect kriging model, especially in cases where the stochastic variability is high. In addition, we further consider the important problem of planning the experimental design by proposing a two stage design approach that systematically balances the allocation of computing resources to new design points and replication numbers in order to reduce the uncertainties and improve the accuracy of the predictions.

Chapter 7 summarizes this studies for the kriging metamodel in stochastic simulation and provides some directions for future research.

Chapter 2

LITERATURE REVIEW

In this chapter, we will provide reviews on several commonly used metamodels first and then focus on the more promising kriging model later on in the first section. In the second section, we review different experimental design methods based on space-filling criterion. In the last section, we look into several metamodel based approaches for simulation optimization.

2.1 Review of Metamodels

Metamodels are built based on the data collected from the target simulation system which can be simplified as the stochastic black box system. The only information available is the combination of the simulation's input sample vector $\boldsymbol{X}$ and output vector $\boldsymbol{Y}$. As a result, the metamodel can be mathematically expressed in Equation (2.1).

$$\hat{f}(\sim) = \hat{f}_{\boldsymbol{X},\boldsymbol{Y};\theta}(\sim) \tag{2.1}$$

where $\hat{f}(\sim)$ is the metamodel, the approximation of the true simulation model, θ is the metamodel's parameters. $\hat{f}(x_0)$ is the output of the metamodel, the prediction of the actual simulation model's outputs $y_0 = f(x_0)$.

2.1.1 Polynomial Regression Model

Polynomial regression model are the most popular and simplest metamodel, as the regression parameters are estimated based on only the simulation model's

input-output combinations: $(\boldsymbol{X}, \boldsymbol{Y})_{\boldsymbol{X}=(x_1,x_2,...,x_n);\boldsymbol{Y}=(y_1,y_2,...,y_n)}$. A typical first-order polynomial regression metamodel will have the form in Equation (2.2).

$$\hat{f}(\boldsymbol{X}) = \sum_{i=1}^{n} \beta_i x_i \tag{2.2}$$

where $\beta_i, i \in [1, 2, ..., n]$ are least square coefficients. The coefficients are selected by minimizing the mean of the sum of squared errors. Generally speaking, the coefficients can be given as in Equation (2.3).

$$\beta = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} \tag{2.3}$$

where $\boldsymbol{X}$ is the observed input vector, and $\boldsymbol{Y}$ is the observed output vector. Polynomial regression model has been well applied in the simulation context. Kleijnen (1998) gave a comprehensive study on the use of polynomial regression model in simulation. In financial engineering, the polynomial regression model has been well applied in the risk analysis and mutual fund evaluation, details can be found in Ruppert (2010). The least square model intends to describe the target simulation model behaviors in the entire sample space with one simple function. This may show inadequacy in terms of the prediction accuracy. For example, in many real world cases, some local behavior might show highly nonlinearity which cannot be captured by a quadratic model, Cheng & Kleijnen (1999a) discussed the use of polynomial regression model in queueing model with highly heteroscedastic responses. Though increasing the degrees of the model could be helpful in some ways, it also would introduce oscillation into the prediction, especially at those locations which are far away from the observations. As a conclusion, least square model is still in common use, but due to its poor prediction capability, it is not a good choice for large-scale or complex system.

2.1.2 Spatial Correlation Model

Spatial correlation metamodel is derived from geo-statistics, which is also known as kriging metamodel. This method assumes that all the points in the sample space are spatial correlated, which means that there are influences between any two points and the intensity of the influence is based on the distance and the distance only.

The key of the kriging metamodel is the assumption of normality. Hence the simulation output $\boldsymbol{Y}$ can be modeled as a Gaussian Random Process (GRP), accounting for the spatial correlated behavior. For instance, the simulation output $y_i = f(x_i)$ at location x_i follows normal distribution $N(\mu_i, \sigma_y^2)$, and the covariance between simulation outputs y_i and y_j can be derived as $Cov[y_i, y_j] = \sigma_y^2 R(y_i, y_j)$. The correlation function $R(y_i, y_j)$ can take varied forms, and the most commonly used is the power exponential family of correlation functions for its smooth response characteristics.

$$R(y_i, y_j) = \prod_{k}^{p} \exp(-\phi_{y,k}(x_{i,k} - x_{j,k})^t) \tag{2.4}$$

As can be seen in Equation (2.4), the correlation function $R(\sim)$ evaluates the spatial correlation between Gaussian random variables y_i and y_j based on the distance between two observations and other controlling parameters.

Structurally, the kriging meta-model's predictor is a linear predictor, which has the following form in Equation (2.5).

$$\hat{f}(x) = \sum_{k=1}^{m} \lambda_k y_k \tag{2.5}$$

where y_k represents the observed simulation output at location x_k, λ_k is the unknown weighted coefficient which is a function of correlation function $R(\sim)$ and observed vectors $\boldsymbol{X}, \boldsymbol{Y}$. Combining Equation (2.4) and Equation (2.5), the kriging predictor is kind of linear predictor of the observed simulation outputs $\boldsymbol{Y}$ and weighted on the spatial correlation and observed data.

Kriging metamodel was first introduced into DACE by Sacks *et al.* (1989). Since then, the metamodel has been widely used in deterministic simulation scenario. The kriging metamodel is the metamodel we propose to use in this research, more detailed introductions and reviews for the application in stochastic simualtion will be given in Section 2.2.

2.1.3 Multivariate Adaptive Regression Splines Model

The multivariate adaptive regression splines (MARS) metamodel is based on simple linear splines model and was introduced by Friedman (1991). It is a

linear model with a forward stepwise algorithm to select model term followed by a backward procedure to prune the model. The general form of the MARS piecewise linear approximation can be given in Equation (2.6).

$$\hat{f}(x) = \beta_0 + \sum_{k=1}^{m} \beta_k B_k(x) \tag{2.6}$$

β is the unknown coefficients and $B_k(x)$ is the basis function which has the form as following form in Equation (2.7).

$$B_m(x) = \prod_{k=1}^{L_m} [S_{i,k}(x_{v(i,k)} - j_{i,k})] \tag{2.7}$$

Here the domain is divided into intervals whose endpoints are called knots. $B_m(x)$ is the linear combination of a series linear functions in different intervals. As the algorithm go forward, the basis function update with the truncated linear function involving a new variable. In Equation (2.7), $x_{v(i,k)}$ is the input variable corresponding to the ith truncated linear function and $j_{i,k}$ is the knot value for $x_{v(i,k)}$. The algorithm will stop when the smoothness of continuity achieves a certain degree.

MARS was further developed by several researchers. Dyn & Yad-Shalom (1991) suggested the optimal distribution for the knots, and McMahon & Franke (1992) minimized the location for the knot points in order to improve the metamodel's performance. Bakin *et al.* (1992) introduced a second order B-splines as the truncated linear function. In simulation application, MARS was compared with several other parametric and nonparametric methods, see Jin *et al.* (2000) and Munoz & Felicisimo (2004). The MARS metamodel outstands for its faster computation in high dimension complicated problems and better estimation accuracy compared to the linear model, principal component regression and classification and regression tree.

2.1.4 Radial Basis Function Model

Radial basis function (RBF) was first introduced by Hardy (1971). The model uses linear combinations of a radially symmetric function based on Euclidean

distance of the form showed in Equation (2.8).

$$\hat{f}(x) = \beta_0 + \sum_{k=1}^{n} \beta_k ||x - x_k|| \tag{2.8}$$

Replace $\hat{f}(x)$ with the observation data vector, then solve the linear formula for the unknown coefficients to obtain the predictor. Meckesheimer *et al.* (2001) and Meckesheimer *et al.* (2002) have applied the new multi-quadratic general form of RBF given in Equation (2.9).

$$\hat{f}(x) = \beta_0 + \sum_{k=1}^{n} \beta_k b||x - x_k|| \tag{2.9}$$

the basis function b can have several choices: linear, cubic, thin plate spline, Gaussian, Inverse multi-quadratic and multi-quadratic. In all these basis functions, the Gaussian and multi-quadratic forms perform best overall, see McDonald *et al.* (2007).

In a recent decade, RBF has been under intensive researches and investigations. It can be treated as a single layer neural network method, which makes it outperform other traditional linear models. Hussain *et al.* (2002a) gave a comparative study on RBF and polynomial regression model as the metamodeling techniques in simulation context. Since the RBF is a mesh-less technique, it was used in the numerical simulation related with Partial Differential Equation (PDE), proposed by Kansa (1990) and widely used in recent few years, see Rocca & Power (2005) and Liu *et al.* (2005). Compared to the popular finite element analysis technique, RBF can fix the ill-posed problem and raise the computational accuracy, which makes RBF sufficient in numerical simulation problems related to complicated unstable engineering circumstances like metal deforming, crystallization process and Micro Electro-Mechanical Systems (MEMS).

2.1.5 Artificial Neural Network Model

The least square model, MARS and Kriging depends on the polynomial equation, either locally or globally. However for Artificial Neural Networks (ANN), the model is divided into 3 different layers: input layer, hidden layer and output

layer. The input from the input layer will be transformed into the nodes in hidden layer. After the recombination of nodes, the output is generated. Since the recombination methods used here refer to some nonlinear optimization skills which is inspired by the mechanism of human nerve system, the model in named as neural networks. It is highly adapted to nonlinear situation and do not require any kind of prior information. As a result, neural network is commonly used in multi-disciplinary research fields, especially good for those extremely complicated or unknown analytical solution, see Law (1994). ANN was first claimed to be capable of providing a satisfied global approximation to any measurable function in Hornik *et al.* (1989) and Funahashi (1989). Kilmer *et al.* (1997) established ANN as a meta-modeling method in discrete stochastic simulation and the model has been widely used in all kinds of simulation scenario, see Anker & Jurs (1992) and Hsu *et al.* (1995). Other models, like kernel smoothing model and Support Vector Machine (SVM) are also the models often used in metamodeling. Compared with all the other metamodels, kriging model can offer outstanding global view of the sample space without losing the local details and it is less time-consuming than other popular AI techniques like ANN, SVM and RBF. Moreover, unlike those non-parametric methods, kriging model owns a clear structure, which makes the interpretation clearer and stronger.

2.2 Review of Kriging Metamodel in Computer Experiments

As stated in Chapter 1, the kriging metamodel was originated in the mining technology, it was later developed and concluded by Matheron (1963). The metamodel has been widely applied in DACE since Sacks *et al.* (1989), and it showed good adaptability and performance in varied real world practices. Welch *et al.* (1990) proposed to use the kriging metamodel in the computer simulation for the Very Large Scale Integrated (VLSI) circuit design in order to reduce the simulation cost. Similar application of the kriging metamodel also can be seen in Gupta *et al.* (2006b) where the metamodel is adopted for its capability and low running cost to fit a sophisticated response surface to the optimal parameter selection

problem of a electronic packing system. Other than the applications in computer simulation models, kriging metamodel also had been applied to other complex systematic problems, like the image process, control system design and so on, see Pham & Wagner (1994) and Wu & Sun (2007). On the other hand, Currin *et al.* (1991) and Morris *et al.* (1993) provided a Bayesian perspective of the kriging model together with numerical examples to provide theoretical analysis and justify the metamodel's capability of estimating the behavior of computer simulation model. For other aspects of the metamodel, Zimmerman & Cressie (1992b) investigate the parameter estimation uncertainty issue for the general Gaussian linear model and concluded that the total prediction error of the metamodel might be inflated by adopting the parameters estimated from the observed data. More details on the metamodel and related discussions can be found in Welch & Sacks (1991), Welch *et al.* (1992) and Santner *et al.* (2003).

The previous studies of the kriging metamodel focus on the deterministic cases, however in recent few years, some researchers started to expand kriging model to stochastic case, especially the stochastic computer simulation model. Beers & Kleijnen (2003) and Kleijnen & Beers (2005) investigated the kriging model's application for the stochastic simulation with constant and non constant variance. The research showed that the performance of existing kriging metamodel varied as the noise pattern change. As a result, the discussion of kriging model's application in stochastic simulation can be further divided into two categories: the homoscedastic case suggesting simulation model with constant variance and the heteroscedastic case suggesting the simulation model with non constant variance. We will separately review these two cases in the following sections.

2.2.1 Kriging Metamodel in Homoscedastic case

As previously mentioned, the kriging metamodel was first developed and applied in the field of geo-statistics. In actual practice of geo-statistics, the data gathered from real world observations usually contains random error. According to Cressie (1993), this random error mainly caused by two factors: micro-scale variation and measurement error. The common practice for modeling this random error is

to assume that the error is normally, independently and identically distributed. The influence of this constant variance noise (or white noise) on the kriging metamodel is usually described as the "nugget effect", which is a term used to describe the discover of gold nugget in mining process, see Cressie (1993). Hence the unknown process with white noise can be modeled by the kriging model with nugget effect (or nugget effect model), see Matheron (1963) and Cressie (1993). The nugget effect model assumes a stationary Gaussian random process for the unknown process, indicating the random error has unknown homogeneous variance. The unknown homogeneous variance is usually estimated by the variogram for the applications in geo-statistic. Cressie (1993) studied the nugget effect model in geo-statistics background and suggested several different variograms in detail, which provided a useful guideline for the kriging model's application in geo-statistics. Other than its application in geo-statistics, the kriging model was introduced into the field of computer simulation by Sacks *et al.* (1989). For the traditional computer simulation models, the simulation outputs are always assumed to be deterministic. For the deterministic case, Santner *et al.* (2003) summarized the characteristics of kriging model and provided a useful general framework for the kriging model's application in deterministic computer simulation context. O'Hagan *et al.* (1999) and Kennedy & O'Hagan (2001) investigate the usage of kriging metamodel in the simulation calibration problem. Other than the deterministic simulation application, recent interest in the stochastic simulation like the Discrete Event Simulation (DES) keeps increasing. Mitchell & Morris (1992) claimed that the kriging metamodel was potentially appropriate for the application with stochastic simulation response. Barton (1992) and Barton (1998) further discussed the possibility that applying kriging metamodel in stochastic simulation model. For the stochastic simulation with homogeneous variance, Kleijnen (2008) suggested that the kriging model with nugget effect still can provide satisfied prediction result. In addition, according to the results provided in Kleijnen & Beers (2005), Sasena *et al.* (2001) and Huang *et al.* (2006), the nugget effect can model the randomness of the simulation output given the stationary Gaussian random process assumption of the simulation model. All these research handled the homoscedastic simulation output with the nugget effect model and suggested that the nugget effect model can provide satisfactory

result in homoscedastic case. According to all the previous studies, the nugget effect model can provide satisfactory results in both the geo-statistics and computer simulation with constant variance noise contexts. However, the methods used to estimate the nugget effect are quite different in these two fields. Sasena *et al.* (2001) claimed that the estimation of the nugget effect value in computer simulation is quite different compared with the variogram used in geo-statistics. The pilot designs or preliminary studies of the simulation model are usually needed in order to provide useful information of the unknown variance of the random error to estimate the nugget effect value. To summarize, the kriging metamodel with nugget effect can provide satisfactory performance in stochastic simulation with homogeneous variance given sufficient information of the simulation model.

2.2.2 Kriging Model in Heteroscedastic case

Other than the homoscedastic case discussed in the previous section, the heteroscedastic case is the more general scenario met in stochastic simulation. According to Kleijnen (2008), the stochastic simulation outputs in practice usually have heterogeneous variance. Therefore, it is worthwhile to investigate the performance of the kriging model in heteroscedastic case. Beers & Kleijnen (2003) and Kleijnen & Beers (2005) looked into the application of nugget effect model in M/M/1 queue with heterogeneous variance and claimed that the nugget effect model could not be directly used to heteroscedastic case without preliminary studies of the simulation model. Yin *et al.* (2008) further investigated the application of the nugget effect model's application in several other heteroscedastic cases. All of these research studied the nugget effect model's performance in heteroscedastic case and concluded that the nugget effect model developed for homoscedastic case was not suitable for heteroscedastic case since it could not handle the non constant variance. The nugget effect model can only be adopted for the heteroscedastic case with certain preliminary studies or pilot designs. The heteroscedastic case can be changed into homoscedastic case with the prior information of the simulation system collected in the preliminary studies, after that the nugget effect model can be applied. However, sufficient computing budget is needed to be appropriately allocated among all the observed points in order

to reduce the variance and change the heteroscedastic case into a homoscedastic case. Beers & Kleijnen (2003) and Kleijnen & Beers (2005) proposed the studentization method which can standardize the stochastic simulation output with extra computing budget. With the standardized simulation outputs, the nugget effect model can provide satisfactory results in the heteroscedastic case. However, the computing budget is limited in many real world situations where the cost of running simulation experiment is unacceptable. The kriging model should therefore be modified before it can be adopted in the heteroscedastic case. In order to improve the kriging model's performance in heteroscedastic case, Yin *et al.* (2008) and Ankenman *et al.* (2010) proposed the kriging model with modified nugget effect and stochastic kriging respectively as more reasonable solutions for kriging model's application in stochastic simulation with heterogeneous variance. The kriging model with modified nugget effect (modified nugget effect model) improved the nugget effect model by changing the assumption on the variance of random error from homogeneous variance to heterogeneous variance. Hence the modified nugget effect model could more appropriately account for the stochastic response with heterogeneous variance than the nugget effect model. This is partially done by penalizing the prediction output at location with high additional variability and compensating the parameter estimation uncertainty caused by the random error in the observed data. The stochastic kriging was developed based on the deterministic kriging model. It considers the additional noise component ε as the intrinsic uncertainty of the simulation itself. Furthermore, Chen *et al.* (2012) go onto look at the effect of Common Random Number (CRN) on the model. The basic assumption for the stochastic kriging model is that the random error can be modeled as an independent stochastic process with zero mean and heterogeneous covariance structure. Based on the numerical results provided in Ankenman *et al.* (2010), the stochastic kriging model outperforms the kriging model with nugget effect and the deterministic kriging model in the heteroscedastic case like the M/M/1 queue example. The modified nugget effect model and stochastic kriging model provides a promising approach to handle the stochastic inputs with heterogeneous variance. However, other issues like the parameter estimation and experimental design still need further investigation for the heteroscedastic case.

2.3 Review of Design of Experiment for Computer Simulation

Experimenters use the design of experiment to increase the information gained from the experiments and decrease the relevant time and cost. The experiments can be divided into two different categories: the physical experiment and the computer experiment. The physical experiment refers to the experiment conducted on the real world physical system. Due to the existence of the random noise such as the measurement error, the result of the physical experiment is usually contained and not repeatable, which increases the difficulty of the data analysis. Statistical methods like the replications and factorization are usually applied to reduce the influence of the random noise and discover the relationship between the inputs and outputs of the system. Unlike the physical experiment, the computer experiment is implemented on the computer simulation model. The inner mechanism of the computer simulation model is dependent on the code of the computer program, which make the outputs of the model is controllable and repeatable. Even for the discrete event simulation with stochastic outputs, the results are still repeatable by controlling the computer coded random number generator. Hence, the design of experiment for computer simulation has several features:

- The deterministic computer experiment does not need replications as it provides fixed outputs for the given inputs. For the stochastic simulation like discrete event simulation, the replication method might be needed to reduce the variability of the data. Hence the experimenter should consider the location of design points and the replications taken at design points at the same time for experimental design in stochastic simulation.

- The computer experiment is usually conducted in the sequential manner for the characteristics of the computer program itself.

- The metamodel is always involved in the design of computer experiment to provide a simplified version of the simulation model for higher computational efficiency and lower running cost.

Considering the design of experiment for computer simulation with metamodel, the main objective of the experimental design together with metamodel is to obtain a set of data for the metamodel in order to provide the best fit of the simulation model or achieve certain design criterion. The designs can be differentiated on the purposes of the experimenter. According to Santner *et al.* (2003), the experimental designs for the computer experiments can be categorized as the space-filling type of design and the criterion-based type of design.

2.3.1 Space-filling Designs

The main purpose of the space-filling designs it to evenly spread out the design points in the sample space to obtain satisfactory estimation of the target simulation model with less bias and lower variation. For the space-filling design, the dispersion of the design points determines the characteristic and performance of the design. Assuming the design $\boldsymbol{D_n}$ contains design points $[x_1, x_2, \ldots, x_n]$, different types of designs can be distinguished from the selection of $\boldsymbol{D_n}$. However, as the space-filling design mainly focus on the allocation of the design points, it can be incorporated with other design methods to better handle the stochastic responses. Especially in the cases with heterogeneous variance, evenly distribution of the computing budget in the whole sample space may be insufficient for the locations with higher variabilities. In such scenarios, the space-filling design can be typically applied as the initial design for the sequential stochastic simulation experiment to provide a rough view of the behavior of simulation model. Given the initial design, other design and allocation methods can be introduced to use the computing budgets more efficiently. In this section we review several most popular space-filling designs in the field.

2.3.1.1 Latin hypercube design

The Latin Hypercube Design / Latin Hypercube Sampling (LHD/LHS) was proposed by McKay *et al.* (1979). It is a randomly generated design, which means that there are multiple equivalent LHD designs for any given number of design points n. For a sample space with input dimension p, the LHD first divide each dimension of the sample space into an equally n intervals, which will results in n^p

equal-sized cells. A selection of n cells are randomly picked from all the possible n^p cells with its projection onto each dimension uniformly distributed among the n intervals. This design is a LHD with n design points and input dimensions p, denoted with $LHD(n, p)$. LHD is commonly adopted by the computer simulation practitioner for its computational simplicity and capability of handle large size data,see Sacks *et al.* (1989) and Fang *et al.* (2006). Other than the original LHD method, the researcher put efforts in improving the LHD. Owen (1992) proposed the randomized orthogonal arrays to improve the projection properties of LHD. LHD can also be combined with other criterions like the IMSE or entropy to provide the optimal LHD method, see Sacks *et al.* (1989) and Shewry & Wynn (1987). To sum up, the LHD shows better capability than simple random sampling method by reducing variation of the sample data and it is relatively easy to realize with computer simulation model. Hence LHD is one of the most popular design methods in computer experiment.

2.3.1.2 Uniform design

Fang (1980) and Wang & Fang (1981) proposed the uniform design as an alternative choice for the space-filling design. Different from the randomly generated LHD, uniform design is a deterministic design. Given the bounded sample space χ, the experimenter can obtain the empirical distribution F for the randomly selected sample $\boldsymbol{X}$. Hence we can define the L_k discrepancy as $D_k = \int_\chi |F_n(x) - F(x)|^k dx$, which evaluates the uniformity of the design. The uniform design can be obtain by minimize the L_k discrepancy function. According to Santner *et al.* (2003), the uniform design can control the absolute error of the desire statistic over χ. Fang *et al.* (2000) claimed that the uniform design can be orthogonal with high possibility, and this is a desirable property for experimental design with computer simulation model. Ma *et al.* (2002) showed that this might be true for many cases but it does not hold for all the cases. Although the uniform design has some desirable properties, its application with the computer experiment still need further investigations.

2.3.1.3 Distance dependent design

The distance dependent design refers to the design that define the dispersion of the design points based on the distance. Johnson *et al.* (1990) introduced the maximin and minimax distance design to uniformly allocate the design points over the entire sample space. Given the sample space χ, the maximin design D indicates that the design intends to maximize the minimum Euclidean distance between any two points $d(x_i, x_j)$: $\max_D \min_{x_i,x_j \in \chi}(x_i, x_j)$. The minimax design D minimizes the maximum distance between any arbitrary selected point x_0 and design point x_i: $\min_D \max_{x_i \in \chi}(x_i, x_0)$. The maximin and minimax designs control the distances between the design points and between any other points and the design points. Hence for the spatial correlated deterministic case, these designs can evenly reduce the prediction uncertainties over the entire sample space χ. Morris & Mitchell (1995) further investigated the characteristics of the designs and proposed to combine the maximin/minimax designs with the simulated annealing method and the LHD to search for the optimum for the simulation model.

2.4 Designs Based on Optimization Criterion

Metamodeling is a good way to represent the inner relationship of input and output of a black-box system. With the appropriate metamodel, we can further study the characteristic of the target black-box system. One of the most useful applications is the optimization. In the following section, several optimization methods which integrate the metamodel into the structure of the optimization algorithm itself will be presented.

2.4.1 Response surface methodology

This model is the application of response surface method (RSM) in simulation. RSM was first developed by Box & Draper (1987), and have been used effectively in many different fields. The basic idea of RSM is to use the traditional least square metamodel to approximate the target model. First, a screening test will

offer some original data, and an empirical model is introduced as the approximation. Second, experiment over a sub-region is carried out. Third, the result of previous experiment is used to decide the search direction for next step, which follow the so-called steepest ascent search method. When the search is close to the optimal, higher order model would be introduced in order to fit the true model.

Equation (2.10) shows the general polynomial form of a response surface model used in the metamodeling.

$$\hat{f}(x,\beta) = \beta_0 + \sum_{k=1}^{m} \beta_k x_k + \sum_{k=1}^{m} \sum_{j>k}^{m} \beta_{kj} x_k x_j + \sum_{k=1}^{m} \beta_{kk} x_k^2 + \cdots + \sum_{k=1}^{m} \beta_{k,k,\ldots,k} x_k^n \quad (2.10)$$

Linear least square estimation is often used to estimate the coefficient based on observation data. RSM cannot offer a global view of the sophisticated design space due to its dependency on least square metamodel which always makes local information lost when focusing on global trend.

2.4.2 Trust region method

Trust region method was proposed by Celis *et al.* (1984), which is as known as restricted step methods. The basic idea of trust region method is to build a quadratic model to fit local objective function within a certain trust region. If there is adequate estimation, the trust region increases, otherwise it will decrease and it goes iteratively.

$$f(x_k + d) \approx \hat{f}(x_k + d) = f(x) + \frac{\partial f}{\partial x_k}^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} H_k d \quad (2.11)$$

Equation (2.11) shows the mathematical model of the trust region method, where d is the next move, H is the Hessian matrix of target function $f(x)$. The algorithm will solve the following objective function in

$$\begin{aligned} \min \quad & \hat{f}_k(x_k + d) \\ \text{subject to} \quad & \|d\| < \delta_k \end{aligned} \quad (2.12)$$

Trust region can be easily modified with other approximation methods other than a traditional second-order polynomial. It can be cooperated with other

global approximation method like kriging in Gano *et al.* (2006) and artificial neural network in Mizutani & Demmel (2003). The key point for applying other meta-model in trust region method is the original method is gradient based and the gradient information is not available for some meta-models like kriging and ANN. Other informatics functions are needed to guide the search. For example, in Gano *et al.* (2006)'s work as previously mentioned, the trust ratio function is deployed instead of the gradient function in Equation (2.13).

$$\rho_n = \frac{f(x_n)_{high} - f(x_n^*)_{high}}{f(x_n)_{scaled} - f(x_n^*)_{scaled}} \tag{2.13}$$

where $f(\sim)_{high}$ and $f(\sim)_{scaled}$ are the penalty function for high-fidelity and scaled low-fidelity functions which are used to show the ratio of high-fidelity approximation to low-fidelity approximation and guide the optimization in multiple fidelity optimization. The problem with trust region method is that it is vulnerable to constraint, which is unfortunately happened for most of the actual engineering cases.

2.4.3 Efficient global optimization

Efficient Global Optimization (EGO) is developed by Jones *et al.* (1998) with its root in Bayesian Global Optimization method which is proposed to overcome the weakness of all the other gradient based algorithms. Some initial sample points will be used to build a statistical model and combined with certain Bayesian analysis function to decide those points to explorer next step. Based on Jones *et al.* (1998), the basic procedure for EGO is:

- 1. Build a initial kriging meta-model of the objective function.
- 2. Use cross validation to ensure that the kriging prediction and measure of uncertainty are satisfactory.
- 3. Find the location that maximizes the Expected Improvement (EI) function. If the maximal EI is sufficiently small, stop.
- 4. Add an evaluation at the location where the EI is maximized. Update the kriging meta-model using the new data point. Go to 3.

Given a kriging meta-model, the key point of the EGO is the EI function which is used to guide the further search in the sequential algorithm. According to Williams et al. (2001), the EI function used in EGO for stochastic simulation has the following form showed in Equation (2.14).

$$\mathrm{E}[I(x)] = \mathrm{E}[\max[\hat{f}(x^*) - \hat{f}(x), 0]] \tag{2.14}$$

where x^* is the current best solution. This EI function has been further developed by Huang *et al.* (2006) by adding an augment terms. EGO appears to be a very promising algorithm and is studied extensively in recent years. In Sasena *et al.* (2001) and Sasena *et al.* (2002), the algorithm was claimed to be sufficient and practical in a series of engineering design problem. And it was further developed by A. Sobester & Keane (2002) with gradient enhanced radial basis function. Also, the Sequential Kriging Optimization (SKO) algorithm introduces by Huang *et al.* (2006) was developed under EGO's structure. This SKO extends its capability to stochastic systems and distinguished itself with several other algorithms. Due to its outstanding performance as a global approximation method, kriging metamodel is often used to represent the current data available in EGO. While in this research, we also focus on kriging metamodel, especially its performance in stochastic situation.

Chapter 3

KRIGING METAMODEL WITH MODIFIED NUGGET EFFECT

3.1 Introduction

Computer simulation is commonly used in industry as a tool to aid in studying the system's characteristics and behaviors. It is especially useful in system optimization problems, where the costs can be greatly reduced by running experiments on the simulation models instead of the real systems. As the complexity of the simulation model increases, the computing cost of running experiments on the simulation model becomes much higher. Metamodels have been applied as simplified approximations to the complex simulation model; see Kleijnen (1987) and Kleijnen (1998). Replacing the simulation model with a metamodel in expensive experiments can increase the efficiency and lower the computing costs. A review of metamodel applications in engineering can be found in Simpson *et al.* (2001) . Among the different types of metamodels available, the spatial correlation model, also known as the kriging model, is one of the more promising metamodels as it is more flexible than regression models and not as complicated and time consuming as artificial intelligence (AI) techniques; see Li *et al.* (2010b) for a comparative study. The kriging model was originally developed in the field of geo-statistics; Matheron (1963). It was first introduced into the DACE by Sacks *et al.* (1989).

Recently, there is an increasing interest in adopting kriging metamodels in industrial engineering problems and applications, related research can be found in Ankenman *et al.* (2010), Huang *et al.* (2006), Sakata *et al.* (2007) and Wang *et al.* (2008).

The kriging model is very suitable for deterministic simulation problems. It is attractive for its interpolating characteristic, providing predictions with the same values as the observations. For example, in Gupta *et al.* (2006a), the kriging metamodel is adopted for its interpolating characteristic. For stochastic simulations where the responses at the same location might vary (for example in a simulation of a queueing system), the interpolation characteristic of kriging models becomes less desirable. In order to model the random fluctuations in stochastic situations, the nugget-effect is introduced. The term "nugget" is borrowed from geo-statistics, referring to the unexpected nugget of gold found in a mining process. According to Cressie (1993), page 127, the nugget effect in geo-statistics is caused by two factors: micro-scale variation and measurement error. In this article, we assume that the system studied can be modeled as an L2-continuous random process, see Cressie (1993), page 112, and hence the nugget effect studied here is purely caused by the random measurement error (or random noise).

The nugget effect in kriging assumes second-order stationarity and is typically used to model white noise effect. Most kriging publications assume that the variance of the random error is homogeneous and the kriging model with nugget-effect is sufficient to solve the problem. However, there are many real world situations where the homoscedastic assumption does not hold. These include queueing systems and networks which can be found in many industrial engineering problems. When applying the homoscedastic kriging model in a heteroscedastic case, the fit can be poor, especially when the sample size is small. We illustrate the noisy applications with the simple function displayed in Figure 3.1. The test function consists of a second-order mean function and a noisy function with step variance.

$$Y(x) = Z(x) + \varepsilon(x) = x^2 + \varepsilon(x) \tag{3.1}$$

where ε indicates the random noise component, with variance $\sigma_\varepsilon^2(x) = 0.083$ when $x \in [-5, 2)$, and $\sigma_\varepsilon^2(x) = 8.3$ when $x \in [2, 5]$. In Figure 3.1, the solid line

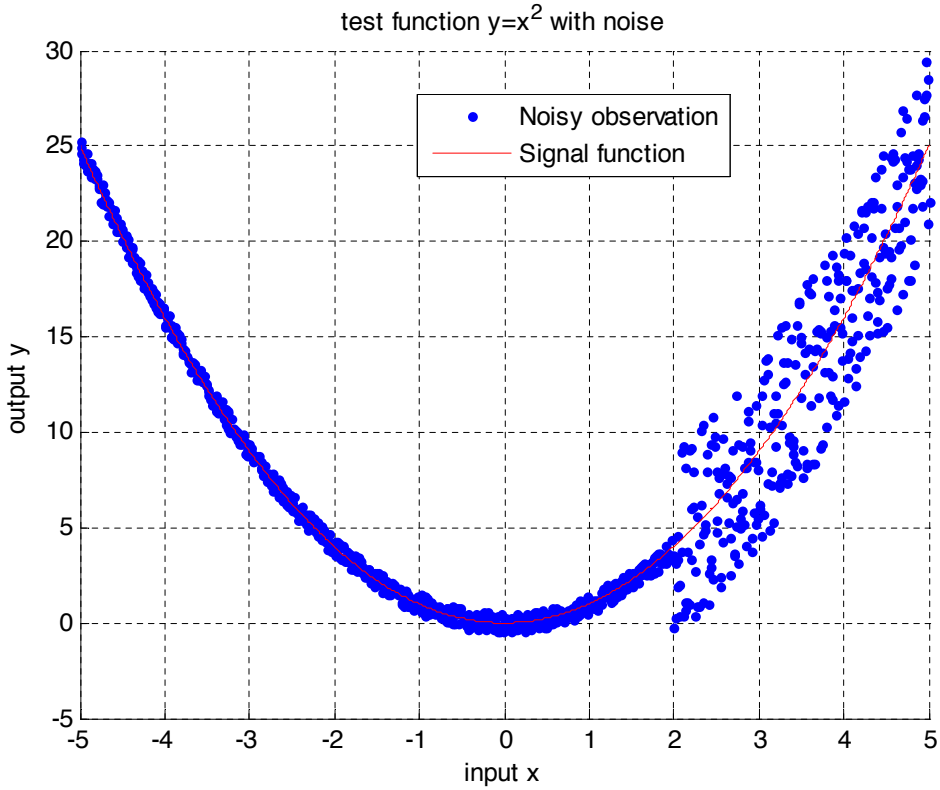


Figure 3.1: Test function with step variance function.

indicates the mean function $Y(x) = x^2$, and the dots are the noisy observations of the mean function $Y(x) = x^2 + \varepsilon(x)$. In the traditional application of kriging in stochastic simulations, replications are taken at each observation point and the averages of the replicates at each point are used as the inputs to the model. Kleijnen (2008), page 92, recommends at least $n \geq 2$ replications to be taken equally at each observation point when no prior knowledge on the variance forms is available, otherwise, the simulation exercise may be meaningless due to the variability in the data. In this test function example, we assume that a budget for only 76 runs is available. Based on this, we spread 19 points from -5 to 5, taking 4 replicates at each point. The averages of the 4 replicates at each of the 19 points are used as the inputs of the model. The solid line in Figure 3.2 plots the fit of the traditional deterministic Ordinary Kriging (OK) model. With limited replications and input points, the ordinary kriging model's predictor output

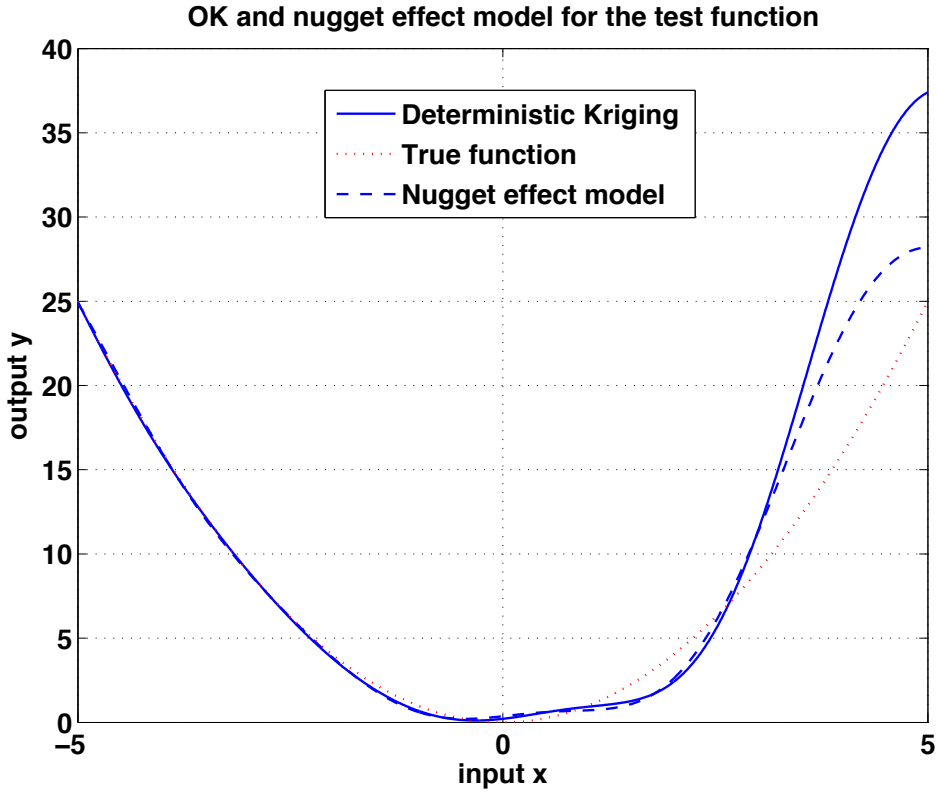


Figure 3.2: Ordinary kriging and nugget-effect model for the test function.

is poor with obvious fluctuations away from the true function when $x < -4$ and $x > 1$. Because the traditional ordinary kriging model is designed under deterministic assumptions, random noise can cause an ill fit and result in disappointing predictions. We note that the predictor output will improve as more replications and observation points are taken. However, in many practical applications of simulation, the computer model can be complicated and time consuming to run, see Gramacy & Lee (2009) and Gupta *et al.* (2006b), limiting the number of observation points and replications that can be taken. Considering the kriging model with nugget-effect which has a homogenous variance assumption, we pool the sample variances at the 19 observation points to estimate the nugget-effect. The predictor output adopting this model is plotted as the dashed line in Figure 3.2. As seen in Figure 3.2, the nugget-effect predictor's output is smoother than the OK predictor. However, in the region $x \in [2, 5]$ where the variance is higher, the

fit is poor compared with the fit in the region $x \in [-5, 2)$. This indicates that the nugget-effect model can still be inadequate as the heterogeneous variance can have an impact on local predictions. Moreover, due to the homogeneous noise assumptions of this model, there is no clear method to estimate the nugget-effect under these heterogeneous conditions. This same phenomenon occurs in the simulation of the M/M/1 queue, one of the most basic queueing models. Beers & Kleijnen (2003) proposed a detrending approach to model out the trend in the data using least squares methods and then apply the deterministic ordinary kriging model to the detrended data. Two alternative methods were later proposed by Kleijnen & Beers (2005) to improve the application of kriging in stochastic problems: the replication method and the studentization method. The replication method proposes that the heteroscedastic problem can be converted into a homoscedastic problem by taking appropriate replications at all the observation locations. This method requires a sequential design with sufficient computing resources to run all the replications. For example, in Figure 3.1, the number of replicates needed in the region with higher variance should be 100 times larger than the number of replicates in the region with lower variance in order to convert the heteroscedastic case into a homoscedastic case. In the study of the M/M/1 queue system for the case where the computing budget is limited, both the OK and nugget-effect model with the application of this replication method can still be inadequate. The studentization method is developed on the basis of the detrended kriging approach. The main idea is to model the trend in the data and then standardize the detrended data. It is an intuitive method to handle inputs with different variances. However, in their numerical examples, this method did not improve much over the OK model. This is due to the amplification of the uncertainty in the estimation of the signal function and variance in the transformation of the predictor, especially when the sample size is small. As seen in Figure 3.2, both the OK model and the nugget-effect model perform poorly when dealing with heteroscedastic data. In this chapter, we relax the stationarity assumption on the covariance process and propose the kriging model with modified nugget-effect to model heteroscedastic observations. This model follows the basic framework of the kriging model with nugget-effect, but extends it by model the random noise $\varepsilon(x)$ as an independent random process from the signal process

$Z(x)$. Moreover, the proposed model takes the sample variance as an additional input to provide variance information. This method has several main benefits: first, the new model retains the original simple structure of the kriging model with nugget-effect; and second, this method allows the independent modeling of the random noise process, which make the estimation of the noise behavior at unobserved location possible; third, computing resources needed for computing the sample variance can be significantly lower than the requirement of the replication method. The sample variance is used as an additional input variable and it can reduce the impact of the heterogeneous variance on the local prediction by penalizing the data with higher variance. In the numerical experiments shown in this chapter, the modified nugget-effect model's performance in the heteroscedastic case is consistently better than the OK model and nugget-effect model.

3.1.1 Differences from the stochastic kriging model

In this chapter, the proposed new kriging model form is based on an extension of the nugget-effect model. Ankenman *et al.* (2010) recently proposed an alternative stochastic kriging model for stochastic simulations. Although the mathematical predictor forms of both models are equivalent (as will be seen in the next Section), our initial assumptions differ in that our proposed modified nugget-effect model is developed from the traditional nugget-effect model, extending it to treat the additional noise component $\varepsilon(x)$ as an independent non stationary random process. The stochastic kriging model is developed based on the deterministic kriging model, and considers the additional noise component $\varepsilon(x)$ as the intrinsic uncertainty of the simulation itself. Ankenman *et al.* (2010) go on to look at the effects of common random numbers on the model and describe experimental design strategies under the stochastic model. In this chapter, our focus differs in that we study in detail the effects and influence of stochastic noise on the traditional deterministic kriging model and nugget-effect model, looking more deeply into the effects on parameter estimation and characteristics of the likelihood function. Unlike the stochastic kriging model, we propose to use the numerical bootstrapping mehtod to estimate the kriging predictor's variance. We

also compare in detail the prediction performances of the three models, providing insights on when each model form is sufficient and adequate.

3.1.2 Organization

This chapter is organized as follows: In the next section, we develop the proposed modified nugget-effect model. We then address the issues of parameter estimation and error measurement and further study the effects of stochastic noise on the traditional models as well as illustrate how the modified nugget-effect model mitigates this problem. In section 3.3, we study the prediction performance and characteristics of the proposed model. Then in section 3.4, the performance of the modified nugget-effect model is illustrated with several numerical experiments and a case study. Comparisons with the traditional kriging model and the nugget-effect model are given, and finally, comparisons with the studentization method are also made.

3.2 Kriging Model with Modified Nugget Effect

In order to introduce the modified nugget-effect model, the details of the kriging model are first discussed. The differences between the modified nugget-effect model, the classic kriging model, and the nugget-effect model will be discussed in three aspects: the development of the modified nugget-effect model, parameter estimation, and error measurement of the model.

3.2.1 Classic kriging (deterministic and nugget effect model)

In Kriging metamodeling, the response of the simulation is treated as a random process $Y(x)$ where x stands for the simulation's p-dimensional ($p \geq 1$) input. Typically, the mean response $Z(x)$ of the random process is of interest. According to Cressie (1993), the general form of the random process can be decomposed in Equation (3.2).

$$Y(x) = Z(x) + \varepsilon(x) = \mu(x) + \delta(x) + \varepsilon(x) \tag{3.2}$$

where $Z(x)$ is the deterministic mean function of the random observations; $\mu(x)$ is the mean of the $Z(\sim)$ function, also known as the large-scale variation; $\delta(x)$ is the bias between the $Z(\sim)$ function and mean $\mu(x)$, also known as the small-scale variation; $\varepsilon(x)$ represents the random measure error (or random noise). In the application of the kriging model to deterministic simulations, the response takes the above form without the random noise component $\varepsilon(x)$. For stochastic simulations with homogenous variances throughout, the nugget-effect kriging model takes the form with $\varepsilon(x) = \varepsilon$. Here we model the random noise as an independent random process. Hence the actual stochastic simulation output $Y(x)$ can be considered as summation of the two independent random processes.

As with most applications of response metamodeling in stochastic simulations, when replicates at each observation point are observed, the sample means of the replicates are typically used as the input for the metamodel estimation. We denote the sample mean and sample variance as:

$$\bar{Y}(x_i) = \sum_{j=1}^{m} \frac{Y_j(x_i)}{n} \tag{3.3}$$

$$s^2(x_i) = \sum_{j=1}^{m} \frac{(Y_j(x_i) - \bar{Y}(x_i))^2}{m-1} \tag{3.4}$$

where $Y_j(x_i)$ denotes the jth replicate at location x_i and m is the number of replications.

Among several kinds of original deterministic kriging models, the one used in this article is the OK model. The OK predictor for point x_0, $\hat{Z}(x_0)$ is a linear combination of all n observation values showed in Equation (3.6).

$$\hat{Z}(x_0) = \sum_{i}^{n} \lambda_i \bar{Y}(x_i) \quad \text{with} \quad \sum_{i}^{n} \lambda_i = 1 \tag{3.5}$$

where λ_i is the MSE optimal kriging weight:

$$\lambda_i = c(x_0)^{\mathrm{T}} R_Z^{-1} e_i + F^{\mathrm{T}} R_Z^{-1} \frac{[1 - F^{\mathrm{T}} R_Z^{-1} c(x_0)]^{\mathrm{T}}}{F^{\mathrm{T}} R_Z^{-1} F} e_i \tag{3.6}$$

$c(x_0) = (corr_Z(d_{01}), corr_Z(d_{02}), \ldots, corr_Z(d_{0n}))$ is the correlation between the point to be estimated and the n observed points, and d_{0i} is the Euclidean dis-

tance between point x_0 and x_i; $corr_Z(\sim)$ represents the spatial correlation between selected two points observed from random process $Z(x)$; R_Z is the matrix of all the spatial correlations between any two observation points; $e_i = [0, 0, \ldots, \underbrace{1}_{\text{the } i\text{th element}}, \ldots, 0, 0]$; and F is the vector of ones with the length of m. It is clear that the kriging weight λ_i is a function of the correlations $corr_Z(\sim)$. Combining Equation (3.5) and Equation (3.6), we can have another form of the OK predictor as shown in Equation (3.7).

$$\hat{Z}(x_0) = F(x_0)\beta_Z + c(x_0)^{\mathrm{T}} R_Z^{-1}(\bar{Y} - F\beta_Z) \tag{3.7}$$

Here in Equation (3.7), β_Z stands for the polynomial regressor, hence the mean function $\mu(x_0) = F(x_0)\beta_Z$. Bias function $\delta(x_0) = c(x_0)^{\mathrm{T}} R_Z^{-1}(\bar{Y} - F\beta_Z)$.

The stationary assumption of the kriging metamodel assumes that the correlation between any two points in the sample space depends only on the distance between the two points. As a result, the covariance function (or its corresponding correlation function) becomes the key component in the model. The general form of the covariance function is given below:

$$v(d_{ij}) = cov(\bar{Y}(x_i), \bar{Y}(x_j)) = \begin{cases} \iota_0(x_i) + \iota_1 & d_{ij} = 0 \\ \iota_1 corr_Z(d_{ij}) & d_{ij} \neq 0 \end{cases} \tag{3.8}$$

where $\iota_0(x_i)$ is the nugget effect value used to describe the variance of the input random noise $\varepsilon(x_i)$, and it can usually be estimated from the sample variance as $\iota_0(x_i) = s^2(x_i)/n$; ι_1 is called the partial sill, representing σ_Z^2, the variance of random process $Z(x)$; $corr_Z(d_{ij})$ is the correlation function based on d_{ij},. So it is clear that the kriging weight λ_i in Equation (3.6) is dependent only on the Euclidean distances between all the observation locations. Further discussion on the covariance function will be given in the following subsections.

For the ordinary kriging predictor, the weights are selected by minimizing the mean squared error defined as:

$$MSE = E[\hat{Z}(x_0) - Z(x_0)]^2 \tag{3.9}$$

The minimization result gives the optimal predictor

$$\hat{Z}(x_0) = \vec{\lambda}\bar{Y} \tag{3.10}$$

where $\vec{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ is the vector of the kriging weights given in Equation (3.6) and $\bar{Y} = (\bar{Y}(x_1), \bar{Y}(x_2), \ldots, \bar{Y}(x_m))$ is the observation vector of sample means, see Cressie (1993), page 123. The minimal mean squared prediction error (also known as the kriging variance) is then given by Cressie (1993), page 123.

$$\begin{aligned} MSE(x_0) &= \sigma_Z^2 \left[1 - \left[c_0 + F \frac{1 - F^{\mathrm{T}} R_Z^{-1} c_0}{F^{\mathrm{T}} R_Z^{-1} F} \right]^{\mathrm{T}} R_Z^{-1} c_0 + \frac{1 - F^{\mathrm{T}} R_Z^{-1} c_0}{F^{\mathrm{T}} R_Z^{-1} F} \right] \\ &= \sigma_Z^2 \left[1 - c_0^{\mathrm{T}} R_Z^{-1} c_0 \frac{(1 - F^{\mathrm{T}} R_Z^{-1} c_0)^2}{F^{\mathrm{T}} R_Z^{-1} F} \right] \end{aligned} \tag{3.11}$$

where c_0 represent the correlation vector $c(x_0)$. As the weights in the kriging predictor are dependent only on the Euclidean distances, this can be inadequate in many heteroscedastic cases where the randomness of the system is also dependent on the location. To solve this problem, the kriging model with modified nugget-effect proposes to relax the stationarity assumption and use the local variance information as an additional input variable. As a result, the predictor can penalize at locations where the variance is high.

3.2.2 The development of kriging metamodel with modified nugget effect

From section 3.2.1, we see that the kriging predictor is a function of the observations $\bar{Y}(x_i)$, $i = 1, 2, \ldots, n$, and the covariance function v. The general form of the covariance function was given in Equation (3.8). Under different underlying assumptions, ι_0 in Equation (3.8) has different forms. We consider two underlying cases: the deterministic case and the stochastic case.

In the deterministic case, the same input at a given location gives the same output. The traditional deterministic kriging model can be used to model this case. In the deterministic kriging model, the random noise is assumed to be 0, so the nugget value ι_0 is 0. The correlation matrix R_Z is given as

$$R_Z = \begin{bmatrix} 1 & corr_Z(d_{12}) & \ldots & corr_Z(d_{1n}) \\ corr_Z(d_{21}) & 1 & \ldots & corr_Z(d_{2n}) \\ \ldots & \ldots & \ddots & \ldots \\ corr_Z(d_{n1}) & corr_Z(d_{n2}) & \ldots & 1 \end{bmatrix} \tag{3.12}$$

and the kriging variance is given in Equation (3.11).

The stochastic case can be further divided into two sub-cases: homoscedastic and heteroscedastic. The nugget effect model is developed under the constant variance assumption to handle the homoscedastic case. As the random noise in this model is assumed to be a constant, the nugget effect ι_0 is a constant which equals the constant variance. The combination of correlation matrix and the nugget effect term is then given by

$$R = \begin{bmatrix} 1+\frac{\iota_0}{\iota_1} & corr_Z(d_{12}) & \dots & corr_Z(d_{1n}) \\ corr_Z(d_{21}) & 1+\frac{\iota_0}{\iota_1} & \dots & corr_Z(d_{2n}) \\ \dots & \dots & \ddots & \dots \\ corr_Z(d_{n1}) & corr_Z(d_{n2}) & \dots & 1+\frac{\iota_0}{\iota_1} \end{bmatrix} \tag{3.13}$$

where ι_0/ι_1 represents the ratio of the variance of the input noise to the process variance. According to Cressie (1993), page 123, the kriging variance is

$$MSE(x_0) = \iota_0 + \iota_1 \left[1 - c(x_0)^{\mathrm{T}} R_Z^{-1} c(x_0) \frac{(1 - F^{\mathrm{T}} R_Z^{-1} c(x_0))^2}{F^{\mathrm{T}} R_Z^{-1} F}\right] \tag{3.14}$$

In the heteroscedastic case, the variances of the random noise are different at different locations. Alternative approaches like the replication method have been proposed to modify the heteroscedastic outputs directly to homoscedastic ones in order to apply the nugget-effect model. These methods, however, have limited applicability under tight computing budget constraints as extra replications are typically needed to drive down the random variability.

In this study, we propose the modified nugget-effect model to address the heteroscedastic case. We relax the stationarity assumption in the homoscedastic model, and assume that the random noise is an independent process of the signal function $Z(x)$. but not identical. As a result, the covariance function is given as:

$$v(d_{ij}) = cov(\bar{Y}(x_i), \bar{Y}(x_j)) = \begin{cases} \iota^*_{x_i} + \iota_1 & d_{ij} = 0 \\ \iota_1 corr_Z(d_{ij}) + \iota_2 corr_\varepsilon(d_{ij}) & d_{ij} \neq 0 \end{cases} \tag{3.15}$$

where $\iota^*_{x_i}$ represents the variance of the input random error at location x_i, and can be estimated from the sample variance and number of replications at location x_i as $\hat{\iota}(x_i)^* = s^2(x_i)/m$. Comparing with Equation (3.8), the constant nugget-effect ι_0 is relaxed to become a variable $\iota^*_{x_i}$, which is dependent on location. ι_1

represents the variance of random process $Z(x)$ and ι_2 represents the variance of random process $\varepsilon(x)$. Moreover, $corr_\varepsilon(\sim)$ is the correlation function for random process $\varepsilon(x)$, which has the similar functional form to the $corr_Z(\sim)$ but with different sensitivity parameter ϕ_ε. Accordingly, the correlation matrix with the modified nugget effect terms becomes

$$R = \begin{bmatrix} 1+\frac{\iota_1^*}{\iota_1} & cf(d_{12}) & \dots & cf(d_{1n}) \\ cf(d_{21}) & 1+\frac{\iota_2^*}{\iota_1} & \dots & cf(d_{2n}) \\ \dots & \dots & \ddots & \dots \\ cf(d_{n1}) & cf(d_{n2}) & \dots & 1+\frac{\iota_n^*}{\iota_1} \end{bmatrix} \tag{3.16}$$

where $cf(d_{ij}) = corr_Z(d_{ij}) + \frac{\iota_2}{\iota_1} corr_\varepsilon(d_{ij})$. As shown in Appendix A, the kriging predictor for this heteroscedastic model is provided in Equation (3.17).

$$\begin{aligned} \hat{Z}(x_0) &= \sum_{i=1}^{m} \lambda_i' \bar{Y}(x_i) \\ &= F(x_0)(\beta_Z + \beta_\varepsilon) + c(x_0)^{\mathrm{T}}(R_Z + R_\varepsilon)^{-1}(\bar{Y} - (\beta_Z + \beta_\varepsilon)) \\ &= F(x_0)\beta_Z + c(x_0)^{\mathrm{T}} R_Z^{-1}(\bar{Y} - \beta_Z) + F(x_0)\beta_\varepsilon + c(x_0)^{\mathrm{T}}\Xi_\varepsilon \end{aligned} \tag{3.17}$$

where λ_i' has the similar structure to the λ_i defined in Equation (3.6) with the correlation matrix R_Z replaced by the R defined in Equation (3.16), here R can be decomposed as $R = R_Z + R_\varepsilon$. Ξ_ε is a positive definite matrix that depends on R_Z and R_ε. The kriging variance can be obtained by substituting ι_i^* for the constant term ι_0 in Equation (3.14) where ι_i^* denotes the variance of the input random error at location x_i.

The kriging variance for MNEK predictor can also be provided as shown in Equation (3.18).

$$MSE(x_0) = \sigma_Z^2 \left[1 - c(x_0)^{\mathrm{T}} R_Z^{-1} c(x_0) \frac{(1 - F^{\mathrm{T}} R_Z^{-1} c(x_0))^2}{F^{\mathrm{T}} R_Z^{-1} F}\right] + \sigma_Z^2 \Omega_\varepsilon \tag{3.18}$$

Here Ω_ε is also a positive definite matrix that depends on R_Z and R_ε, which represents the influence of random noise on the predictor variance and will equal to zero when $\varepsilon(x) = 0$.

As discussed, the kriging model with nugget-effect is used to handle the homoscedastic case by adding a constant onto the diagonal of the correlation matrix

of the deterministic kriging model. The model proposed for the heteroscedastic case has a form similar to the kriging model with nugget-effect, with a variable term added onto the diagonal instead. We can rewrite the correlation matrix as the summation of the deterministic correlation matrix R_Z for the deterministic kriging model and an additional matrix R_ε. For the homoscedastic case, R_ε is a matrix with a constant $\eta_c = \iota_0/\iota_1$ on the diagonal:

$$R = R_Z + R_\varepsilon \text{ with } R_\varepsilon = \begin{bmatrix} \eta_c & 0 & \dots & 0 \\ 0 & \eta_c & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \eta_c \end{bmatrix} \tag{3.19}$$

For the heteroscedastic case, η is a matrix with a variable $\eta_i = \iota^*_{x_i}/\iota_1$ on the diagonal:

$$\begin{aligned} R &= R_Z + R_\varepsilon \text{ with} \\ R_\varepsilon &= \begin{bmatrix} \eta_1 & \frac{\iota_2}{\iota_1} corr_\varepsilon(d_{12}) & \dots & \frac{\iota_2}{\iota_1} corr_\varepsilon(d_{1n}) \\ \frac{\iota_2}{\iota_1} corr_\varepsilon(d_{21}) & \eta_2 & \dots & \frac{\iota_2}{\iota_1} corr_\varepsilon(d_{2n}) \\ \dots & \dots & \ddots & \dots \\ \frac{\iota_2}{\iota_1} corr_\varepsilon(d_{n1}) & \frac{\iota_2}{\iota_1} corr_\varepsilon(d_{n2}) & \dots & \eta_n \end{bmatrix} \end{aligned} \tag{3.20}$$

Hence, we call this model the kriging model with modified nugget-effect. Note that substituting R in Equation (3.16) as R_Z in the deterministic predictor in Equation (3.5) gives the same mathematical predictor form as Ankenman *et al.* (2010). Moreover, modeling the random noise $\varepsilon(x)$ as an independent random process allows the experimenter to estimate the noise variance at unknown location, we will discuss several alternative estimation approach in the following section. However, for the ease of illustration, during the discussion on the nugget-effect's influence on parameter estimation in the next section, we assume that the observations from random noise process $\varepsilon(x)$ are independent but not identical. This assumption eases the computational problem by changing the η matrix in Equation (3.20) into a diagonal matrix.

3.2.3 Parameter estimation and characteristics of likelihood function with noisy data

The correlation functions $corr_Z(\sim)$ and $corr_\varepsilon(\sim)$ in R can have different forms. The Gaussian correlation function has been widely applied. It has the following form:

$$corr(d_{ij}) = \prod_{k=1}^{p} \exp(-\phi_k d_{ij}^2), \quad \phi_k > 0 \tag{3.21}$$

where ϕ_k is the sensitivity parameter for the kth input, which is usually estimated by the Maximum Likelihood (ML) approach in the DACE toolbox. This is a widely used free Matlab kriging toolbox that is well documented in Lophaven *et al.* (2002). As for the following discussions in this chapter, we consider the random noise to be independent but not identical. Hence we only discuss the parameter estimation problem for the process $Z(x)$, which can be explained as the noisy data's influence on the estimation of the signal process $Z(x)$.

For stochastic responses, especially in the heteroscedastic case, the likelihood function is likely to have an erratic behavior caused by the high variability of the data. In this case, the estimation of $\phi_Z = (\phi_{Z1}, \phi_{Z2}, \ldots, \phi_{Zp})$ may have a large variance because the likelihood function can be flat near the optimum. Because ϕ_Z is the sensitivity parameter of the correlation function, inaccurate estimation may cause oscillations and fluctuations in the kriging predictor's output. The test function in Equation (3.1) will be used as an example to demonstrate this phenomenon.

First, we will show how the likelihood function for the deterministic model will change with noisy observations. According to Cressie (1993), page 92, the likelihood function for the deterministic kriging model is:

$$\ell(\phi_Z) = \frac{1}{2}\ln\det(R_Z) + \frac{1}{2}(\bar{Y} - F\mu)^{\mathrm{T}} R_Z^{-1}(\bar{Y} - F\mu) \tag{3.22}$$

where R_Z is the deterministic correlation matrix, which is a function of ϕ_Z. The correlation function used here is the Gaussian correlation function, given in Equation (3.21). Obviously, if any $\phi_{Zk} \to \infty$ (for $k = 1, 2, \ldots, p$), $R_Z \to I$(identity matrix), and so, $\lim_{\phi_{Zk}\to\infty} \ell(\phi_Z) \to \frac{m}{2} var(\bar{Y})$. For the one dimensional test function of the form in Equation (3.1) where the variability in the observations is

high, the likelihood function will increase in the region where $\phi_Z \to \infty$. This phenomenon will cause a bad estimate of ϕ_Z. The plot of the likelihood function for both the signal function and the noisy observations on the signal function is shown in Figure 3.3. In Figure 3.3, considering the likelihood function for the

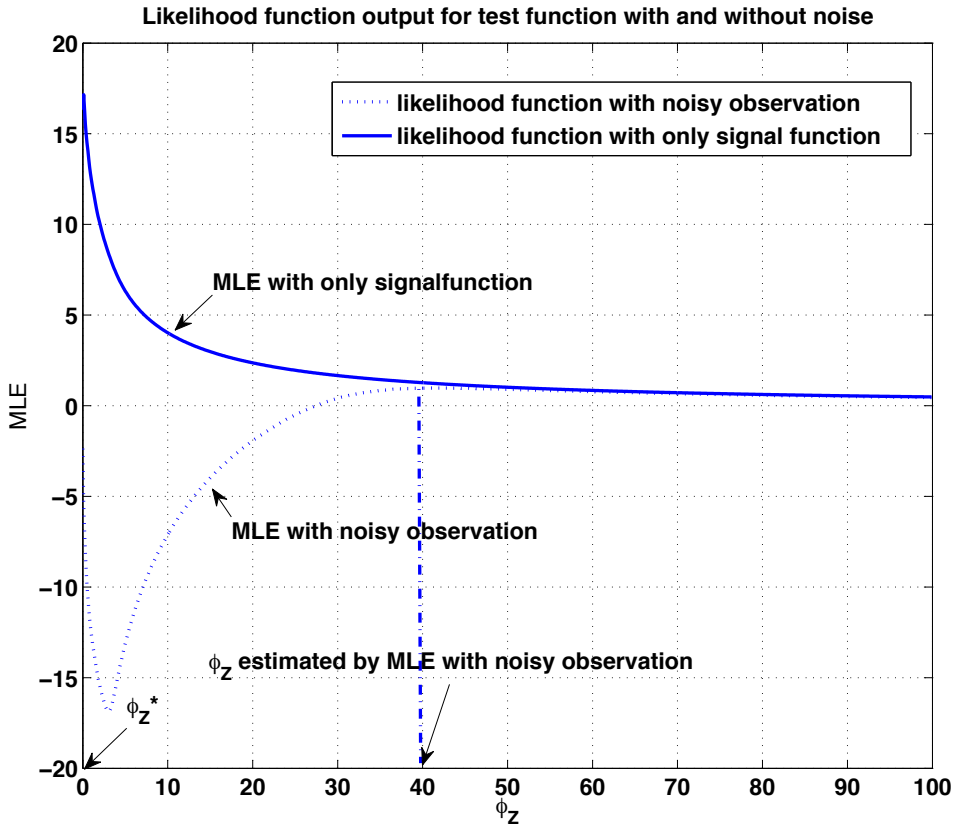


Figure 3.3: Likelihood function for ϕ_Z (signal function only and noisy observation in Equation (3.1)).

signal (solid line), the maximum likelihood estimator (MLE) ϕ_Z^* is very close to 0. ϕ_Z^* can be viewed as the best ϕ_Z possible without the influence of noise. From Equation (3.21), we see that ϕ_Z is the sensitivity parameter of the correlation function. Intuitively, a small ϕ_Z implies that the correlation is not very sensitive to distance, providing a smooth predictor. However, with noisy observations (see the dotted line in Figure 3.3), much of the likelihood function has been lifted up

in the right side of the plot when ϕ_Z gets large. This causes an ML estimated ϕ_Z to be much larger than ϕ_Z^*, making the correlation very sensitive to distance. As a result, the output of the kriging predictor will oscillate. Moreover, as noise gets larger, the variability of the MLE increases. When assuming the kriging model with nugget-effect or modified nugget-effect, this situation is improved. As seen in Figure 3.4, the lifted up portion is pulled back and the likelihood function is corrected with a much smaller ML estimate of ϕ_Z. We can see this from the

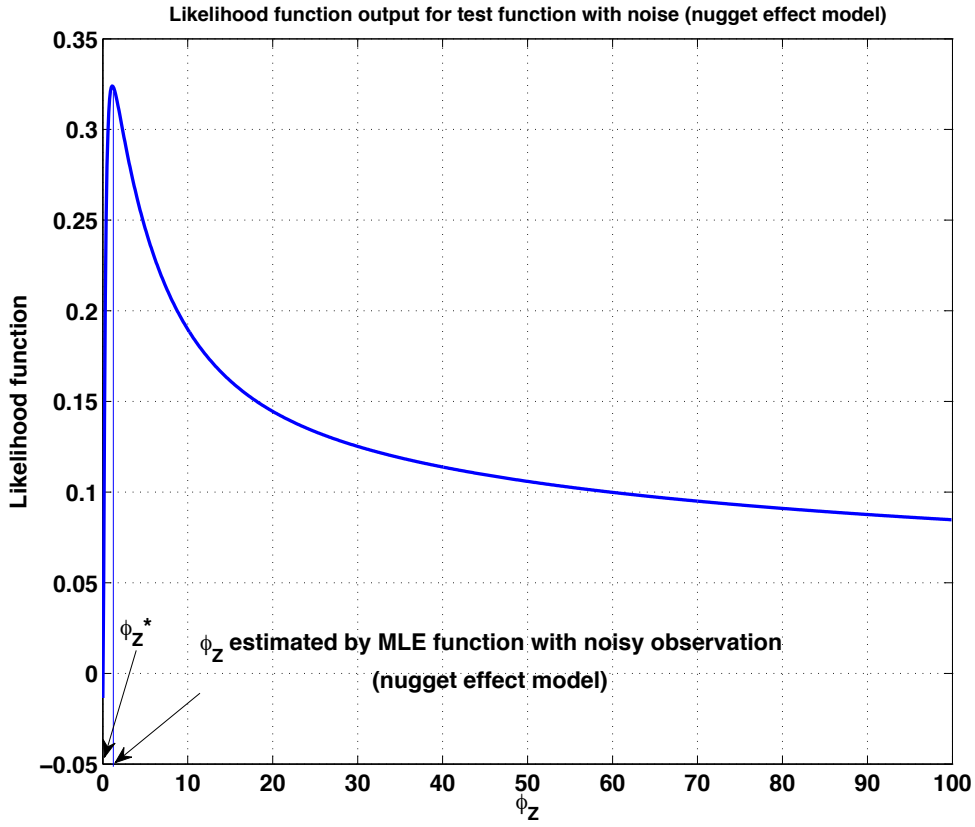


Figure 3.4: Likelihood function for ϕ_Z with nugget effect model (noisy observation of the signal function).

likelihood function of the nugget-effect and modified nugget-effect model:

$$\ell'(\phi_Z) = \frac{1}{2}\ln\det(R') + \frac{1}{2}(\bar{Y} - F\mu)^{\mathrm{T}} R'^{-1}(\bar{Y} - F\mu) \tag{3.23}$$

where $R = R_Z + R_\varepsilon$ and R_ε is given in Equation (3.19) and Equation (3.20) for the nugget-effect and modified nugget-effect models. Because matrix R is positive definite, by the Woodbury identity, see Woodbury (1950), we have

$$R^{-1} = R_Z^{-1} - R_Z^{-1}(R_Z^{-1} + R_\varepsilon^{-1})^{-1}R_Z^{-1}.$$

So, the likelihood function becomes

$$\begin{aligned} \ell'(\phi_Z) &= \frac{1}{2}\ln\det(R) + \frac{1}{2}(\bar{Y} - F\mu)^{\mathrm{T}} R_Z^{-1}(\bar{Y} - F\mu) \\ &- \frac{1}{2}(\bar{Y} - F\mu)^{\mathrm{T}} R_Z^{-1}(R_Z^{-1} + R_\varepsilon^{-1})^{-1}R_Z^{-1}(\bar{Y} - F\mu) \end{aligned}$$

The last term on the right hand side

$$P_\eta(\phi_Z) = -\frac{1}{2}(\bar{Y} - F\mu)^{\mathrm{T}} R_Z^{-1}(R_Z^{-1} + R_\varepsilon^{-1})^{-1}R_Z^{-1}(\bar{Y} - F\mu) \tag{3.24}$$

behaves like a penalty function when ϕ_Z gets large. The plot of $P_\eta(\phi_Z)$ when η defined by Equation (3.19) is given in Figure 3.5. We see that $P_\eta(\phi_Z)$ behaves much Like the penalty function approach proposed in Li & Sudjianto (2005) to correct highly variable likelihood estimates due to the lack of data in computationally intensive deterministic simulation models. From the deterministic perspective, $P_\eta(\phi_Z)$ serves as a natural penalty function to the log likelihood function to reduce the variability of the estimated ϕ_Z when "noisy" observations are obtained, hence reducing the erratic behavior of the predictor. When applying the kriging model with nugget-effect in a heteroscedastic situation, there is no straightforward interpretation for selecting an appropriate nugget value. Kleijnen & Beers (2005) noted that variogram results are meaningless for heteroscedastic data. An ad hoc approach is to pool sample variances as done in the test function example. Here, we provide a more intuitive argument for selecting the nugget-effect value for this model. For cases like the test function in Equation (3.1) with a step variance function, the variance of the random noise is considerably large in some regions and cannot be ignored, increasing the variance of all the observations, $var(\bar{Y})$. For the likelihood function of the nugget-effect model, when any $\phi_{Zk} \to \infty$ (for $k = 1, 2, \ldots, p$), $R \to I$, the function will become:

$$\lim_{\phi_{Zk}\to\infty} \ell'(\phi_Z) \to \frac{m}{2}\ln(1+\eta_0) + \frac{m}{2(1+\eta_0)var(\bar{Y})} \tag{3.25}$$

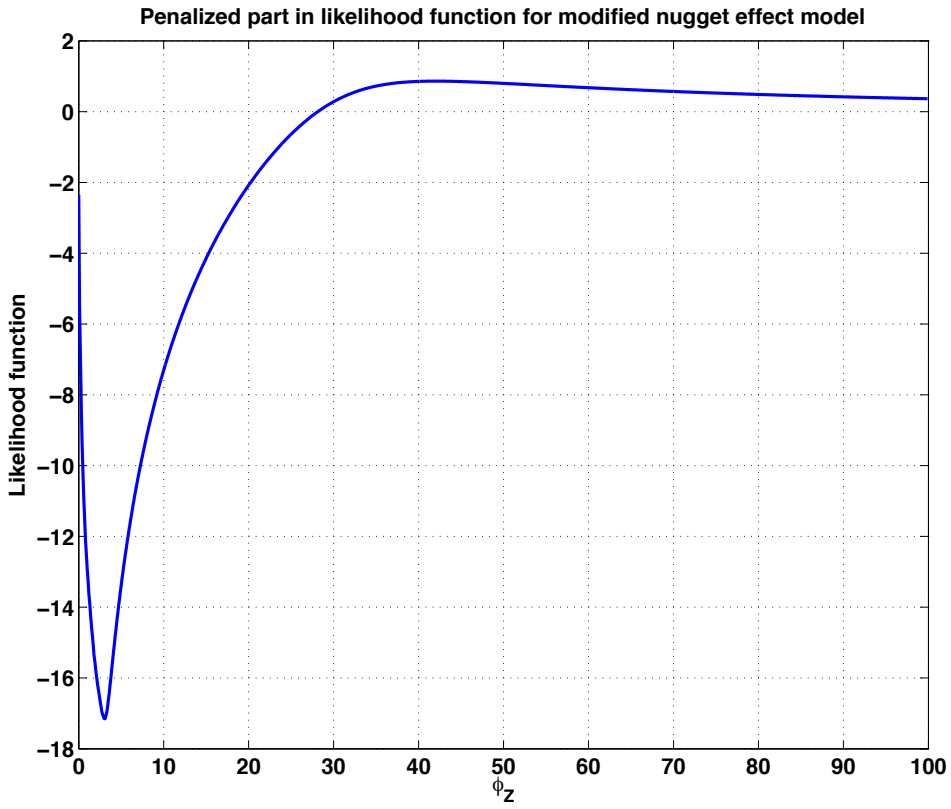


Figure 3.5: Profile of the penalized portion of the likelihood function for modified nugget effect model.

In order to fix the abnormal likelihood function caused by the high variability in the observations, the influence of $var(\bar{Y})$ can be reduced by increasing η_0. To do so, the largest of the heterogeneous variances should be used as the nugget-effect value for kriging metamodel with nugget-effect. We adopt this value in the examples given in section 3.4. The likelihood function for the nugget-effect model in Figure 3.4 has a similar profile as the plot of the likelihood function with only the signal function in Figure 3.3 (the solid line). As a result, the estimation of ϕ_Z by maximizing the likelihood function with nugget-effect will be closer to ϕ_Z^* than the estimation given by maximizing the likelihood function without the nugget-effect. The likelihood function for the modified nugget-effect model has similar behavior and characteristics. Thus, in the heteroscedastic case, both the kriging

model with nugget-effect and the kriging model with modified nugget-effect can provide a better estimator of ϕ_Z than the traditional deterministic kriging model. Therefore, the performance of either kriging model with nugget-effect or kriging model with modified nugget-effect will be much better than the original kriging model, especially when the variability of the data is high, and the number of available observations is small.

3.2.4 Error measurement

In order to compare the performance of the nugget-effect model and the modified nugget-effect model in the heteroscedastic case, an error measurement standard is needed. We use the Mean Squared Error (MSE),

$$MSE(x_0) = \sigma^2_{kg}(x_0) = \mathrm{E}\left[\hat{Y}_{x_0} - Y(x_0)\right]^2 \tag{3.26}$$

The squared error has several benefits: it simplifies the Bayes risk, it has symmetric confidence intervals, and it is related to the variance terms, which is especially suitable for kriging. As can be seen, the MSE is used to measure the difference between the predictor and the observation. In the deterministic case, because the observation strictly equals the mean function, MSE can offer a clear view of the prediction accuracy with respect to the mean function. For the stochastic case however, when predicting at a new location x_0, our interest is in the actual mean function, $Z(x_0)$, and not the observation $Y(x_0)$ which is distorted by random noise. As a result, the kriging predictor will be:

$$\hat{Z}\left(x_0\right) = \sum_{i=1}^{n} \xi_i \bar{Y}(x_i) \tag{3.27}$$

with the kriging weight ξ_i, which is generated by minimizing the mean squared error with respect to mean function $Z(x)$ (also see Equation (3.5)). The predictor is still based on all the observations, $\bar{Y}(x_i), i = 1, 2, \ldots, m$. The notation MSE_S is used for the mean squared error computed with respect to the mean function $Z(x)$, which according to Cressie (1993), page 128, is

$$\begin{aligned} MSE_S(x_0) &= \varsigma^2_{kg}(x_0) = \mathrm{E}\left[\hat{Z}\left(x_0\right) - Z(x_0)\right]^2 \\ &= \iota_1 \left(1 - \left[c + F\frac{1 - F^{\mathrm{T}}R^{-1}c}{F^{\mathrm{T}}R^{-1}F}\right]^{\mathrm{T}} R^{-1}c + \frac{1 - F^{\mathrm{T}}R^{-1}c}{F^{\mathrm{T}}R^{-1}F}\right) - \sigma^2_{\varepsilon}(x_0) \end{aligned} \tag{3.28}$$

where for the nugget-effect model, $\sigma_\varepsilon^2(x_0) = \iota_0$; for the modified nugget-effect model, $\sigma_\varepsilon^2(x_0) = \iota_0^*$. Clearly, the MSE_S can offer the evaluation of the predictor's performance with respect to the mean function in the presence of random noise. In the heteroscedastic case, however, one limitation is that the variance of the random error at every prediction point is needed. As it is assumed that the random error is independent but not identical, the variance information is not available unless the location is observed. Hence, comparing the performances of the kriging metamodel with nugget-effect and the kriging metamodel with modified nugget-effect is not straightforward. As an alternative, we decompose the sample space into two parts: observation points and the areas in between the observations. The MSE_S at all the observation points can be compared, which will be given in the following section. For the areas in between the observations, the prediction is dependent on the prediction at the observations and the sensitivity parameter ϕ_Z. For comparison purposes, we fix the sensitivity parameter ϕ_Z in both models to be the MLE estimated with nugget-effect.

3.3 Prediction Performance of the Kriging Model with Modified Nugget-effect

The prediction performance of a kriging model with modified nugget-effect can be divided into two parts: predictor's output and variance of the predictor.

3.3.1 Comparison through MSE_S

From Equation (3.28), it can be shown that the MSE_S at the ith observation point for the kriging model with modified nugget-effect is

$$MSE_S(x_i) = \iota_1(\eta_i^2 \frac{\Delta_{mi}^2}{\Delta_m} - \eta_i^2 \Delta_{mi}) \tag{3.29}$$

where $\Delta_m = F^{\mathrm{T}} R^{-1} F$ indicates the summation of all the elements in the inverse correlation matrix, Δ_{mi} represents the summation of the ith column or row, and as showed in Equation (3.20), $\eta_i = \iota_i^*/\iota_1$. The details of this derivation are

provided in Appendix B. Similarly, for the kriging metamodel with nugget-effect, the MSE_S at the ith observation point is

$$MSE_S(x_i) = \iota_1(\eta_c^2 \frac{\Delta_{ni}^2}{\Delta_n} - \eta_c^2 \Delta_{ni}) \tag{3.30}$$

Because the difference between Δ_m and Δ_n is typically not significant, assuming $\Delta_m \approx \Delta_n = \Delta_0$ and $\Delta_{mi} \approx \Delta_{ni} = \Delta_{0i}$, when the nugget value ι_0 exceeds $l_0 = \sum_{i=1}^{m} \left[\iota_i^{*2} \left(\frac{\Delta_{0i}^2}{\Delta_0} - \Delta_{0i} \right) \right] / \sum_{i=1}^{m} \left(\frac{\Delta_{0i}^2}{\Delta_0} - \Delta_{0i} \right)$, the modified nugget-effect model outperforms the nugget-effect model in terms of MSE. For most situations, in order to have a better estimation of the ϕ, a larger nugget-effect is desired, as seen in Equation (3.25). Hence, it is often likely that the selected ι_0 will be larger than the l_0.

3.3.2 Estimating predictor's variance

The covariance function of the proposed modified nugget-effect model is given in Equation (3.15). Based on the model assumptions in Section 3.2.1, the random noise can be modeled as an random process independent of the signal process $Z(x)$. Hence it is possible to separately model the variance of the random noise with different methods. One possible method is to use another deterministic kriging model to estimate the variance of the random noise $\varepsilon(x)$, which is also adopted in Ankenman *et al.* (2010). Other numerical approaches can also be used to approximate of the variance of the predictor. Here we propose three different methods to estimate the variance of the predictor.

If we consider the random noise $\varepsilon(x)$ is a random process, then the variance of the random noise can also be treated as an unknown function. In the stochastic simulation with heterogeneous variance, this variance function is location dependent. As the sample variance can be obtained in Equation (3.4), we can assume that the variance of random noise $\sigma_\varepsilon^2(x)$ is a Gaussian Random Process and hence the variance at unknown locations can be estimated by building a kriging model given the observed sample variances. Modeling the variance function as a GRP allows the experimenter to use the local variance information to estimate the variances at the unknown locations and increases the estimation accuracy.

The second method is the nonparametric bootstrap approach. Bootstrapping is a re-sampling method developed by Efron (1979), and has been widely used in estimating the distribution and the properties of estimators. Nonparametric bootstrapping draws samples from the original observed data with replacement to form the bootstrap samples. The bootstrapping statistic is then estimated from the bootstrap samples. Parametric bootstrapping can also be applied as an alternative approach to estimate the variance of the kriging predictor. In the parametric approach, based on the model assumptions in Equation (3.2), bootstrapped samples can be drawn from the normal distribution with plug-in parameter estimators of $\hat{\beta}, \hat{\sigma_Z}^2, \hat{\phi}$; see den Hertog *et al.* (2005). Both these bootstrapping approaches can provide estimators of the predictor variance. However, to avoid the assumption of the plug-in estimator and its additional variability due to the random noise (as discussed in Section 3.3) in the parametric approach, here we apply the nonparametric approach to estimate the variance of the predictor.

Applying the nonparametric bootstrap approach, the bootstrap estimator of the predictor's variance can be given as:

$$\hat{var}(\hat{Z}(x_0)) = \hat{var}^B(\hat{Z}(x_0)) = \sum_{j=1}^{b} \frac{(P_j^B(Z(x_0)) - \bar{P}^B(Z(x_0)))^2}{b-1} \tag{3.31}$$

where $P_j^B(Zx_0)$ denotes the bootstrapped predictor's output at location x_0 with the jth set of bootstrap samples, $\bar{P}^B(Z(x_0))$ is the average of all the bootstrapped predictor's outputs at location x_0 given b sets of bootstrap samples, where b is the total number of the bootstrap sample sets. Given the simulation observations $\bar{Y}(x_i) = \left[\bar{Y}_1(x_i), \bar{Y}_2(x_i), \ldots, \bar{Y}_n(x_i)\right]_{i=1,2,\ldots,n}$, the general bootstrap procedure to estimate the predictor's variance at a new point is given as follows:

Step 1 :For each observation location x_i, re-sample n observations (with replacement) from the n original simulation observations obtained at that location. Repeat this b times to obtain b bootstrap samples of n re-sampled observations at each location x_i, we have

$$\vec{Y}^B(x_i)_j = [Y_1^B(x_i), Y_2^B(x_i), \ldots, Y_n^B(x_i)]_{i=1,2,\ldots,n;j=1,2,\ldots,b};$$

Step 2 :For each bootstrap sample j, $j = 1, 2, \ldots, b$, compute the sample mean $\bar{Y} = \frac{\sum_{k=1}^{n} Y_k^B(x_i)_j}{m}$ and sample variance $\hat{var}(\bar{Y}^B(x_i)_j) = \frac{\sum_{k=1}^{m}(Y_k^B(x_i)_j - \bar{Y}^B(x_i)_j)^2}{m-1}$ at each location x_i. Using these as inputs to the kriging model, we can obtain the bootstrap predictor at location x_0 for the jth bootstrap sample, $P_j^B(Z(x_0))$;

Step 3 :With b bootstrap predictors at location x_0, $P_j^B(Z(x_0))_{j=1,2,\ldots,b}$, the non-parametric bootstrapping predictor variance at x_0 can be computed with Equation (3.29).

For an alternative method, den Hertog *et al.* (2005) recommended that the variance of the predictor can be estimated by interpolation. Because the variances at all the observation points are available, piecewise linear interpolation can be used to cover the space between any two observation points. In the following examples, we compare the kriging model estimated variance, bootstrapping variance and interpolating variance with the empirical brute-force predictor variance generated through 10,000 replications of the independent simulation results.

From Table 3.1, we can see that all the estimated variances are close to the empirical variance of the predictor. The bootstrapping method gives a result slightly higher than the other 3 methods when $x \leq 0.8$. All the three methods give results lower than the empirical method when x approaches to 0.9. For the

Table 3.1: Estimated variance for modified nugget effect predictor for M/M/1 example.

Estimation methods	$x = 0.6$	$x = 0.7$	$x = 0.8$	$x = 0.9$
empirical	1E-3	7E-3	0.082	2.511
kriging	2E-3	9E-3	0.091	2.219
bootstrapping	4E-3	0.048	0.123	2.152
interpolating	1E-3	7E-3	0.089	2.364

M/M/1 model, the variance of the output goes to infinity when the input traffic rate x goes to 1. This phenomenon is also known as the "variance explosion".

The kriging modeling and interpolation methods are much easier and faster than the bootstrapping method. However, as indicated in this study, the kriging modeling estimation suffers from the parameter estimation uncertainty, underlying mean model selection uncertainty and so on. Hence the usage of kriging modeling might be limited in certain situation with limited data although it can provide smoother estimation result than the linear piece-wise interpolation. For the cases with limited observations, the bootstrapping method is preferred as the precision of result with any interpolation methods may be low.

3.4 Examples

In this section, two numerical examples and a short case study will be presented. All the examples used here have heteroscedastic variances: a test function with step variance function, the M/M/1 queueing system and a queueing network problem. The quadratic test function provides insight into the heteroscedastic case, focusing on the comparison between a low and high step variance. The M/M/1 queue is a classic queueing system where analytical results are available for comparison. It has a continuous variance function and is a basic component of many more complex queueing systems. The final case study is of a comprehensive queueing network system where the underlying functional forms are unknown.

3.4.1 Test Function

The test function here is given in Equation (3.1) with a step variance function. In order to illustrate the influence of the heterogeneous variance, we set 3 different ratio levels for the step variance function: $\sigma_\varepsilon^2(x) = 0.083$ when $x \in [-5, 2)$, and $\sigma_\varepsilon^2(x) = 0.83, 8.3, 83$ when $x \in [2, 5]$. Figures 3.8-3.10 illustrate the predictors' outputs (ordinary kriging, kriging with nugget-effect and kriging with modified nugget-effect) for r_{var} (ratio of max variance to min variance) = 10,100, 1000 respectively.

From these three figures, we can see the erratic behavior in the output of the ordinary kriging predictor. This is the result of the high variance in parameter estimation. The output of the kriging predictor with nugget-effect is much

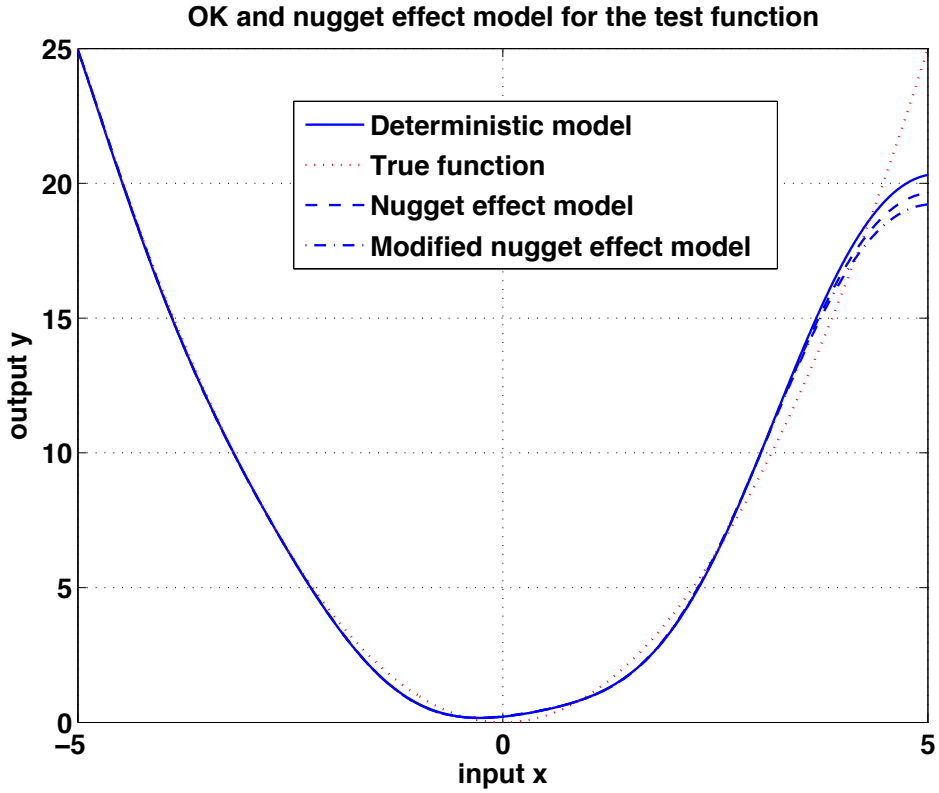


Figure 3.6: Different predictors' output for test function (r_{var} =10).

smoother, but it is still far from the signal function in the region with higher variance, especially for the case when r_{var}=1000. The modified nugget-effect predictor lies closest to the signal function. Selecting a second-order polynomial function ("universal kriging") instead of a constant as the process mean μ in Equation (3.2) can improve the prediction accuracy. Figure 3.9 illustrates the predictor output when a second-order model is assumed in place of the constant model for the three kriging model forms. As seen in this figure, the modified nugget-effect model with a second-order polynomial regression model provides a smoother output over the entire region. Table 3.2 summarizes four error measures MSE_{SO}, MSE_S, AAE_S and MAE_S for the different models and variance ratios used. For the nugget-effect model in Table 3.2, the min variance refers to the nugget value ι_0 equals to the minimum value for the step variance function, and the max variance refers to the nugget value ι_0 equals to the maximum value

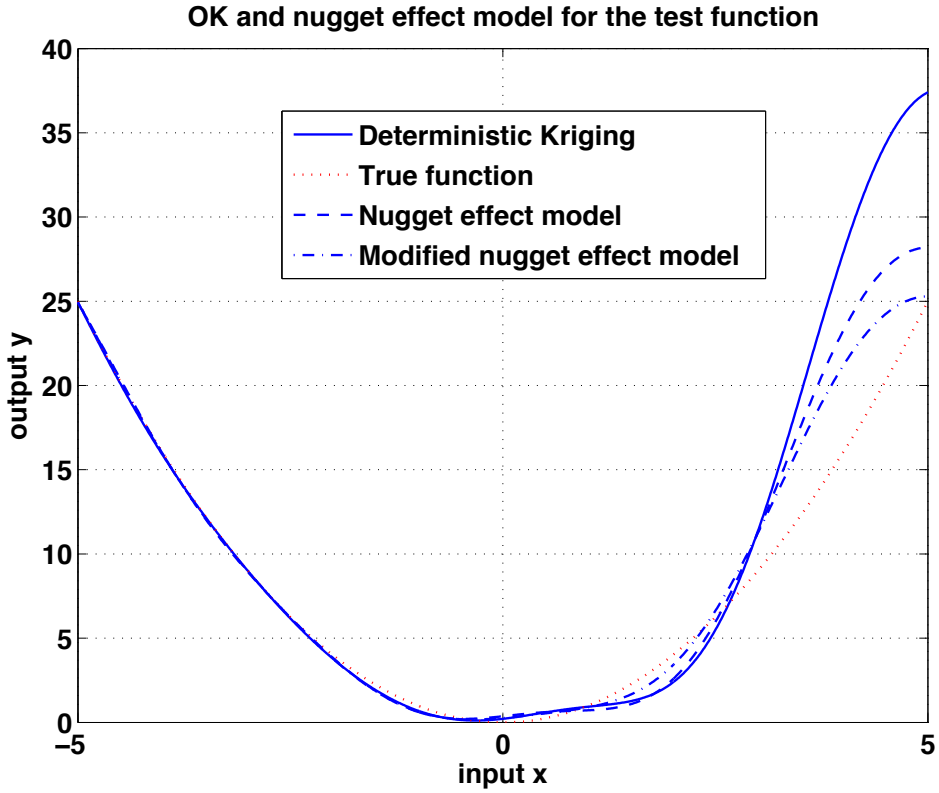


Figure 3.7: Different predictors' output for test function (r_{var} =100).

for the step variance function. The modified nugget-effect (optimal) refers to the modified nugget effect model with the optimal sensitivity parameter ϕ_Z^*. The first measure MSE_{SO} refers to the mean squared error between the predictor's output and signal function at all the observed points:

$$MSE_{SO} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Z}\,(x_i) - Z(x_i))^2 \tag{3.32}$$

The second MSE_S refers to the differences at all the points (including the n=19 observed points and k=982 unobserved test points):

$$MSE_S = \frac{1}{n+k}\sum_{i=1}^{n+k}(\hat{Z}\,(x_i) - Z(x_i))^2 \tag{3.33}$$

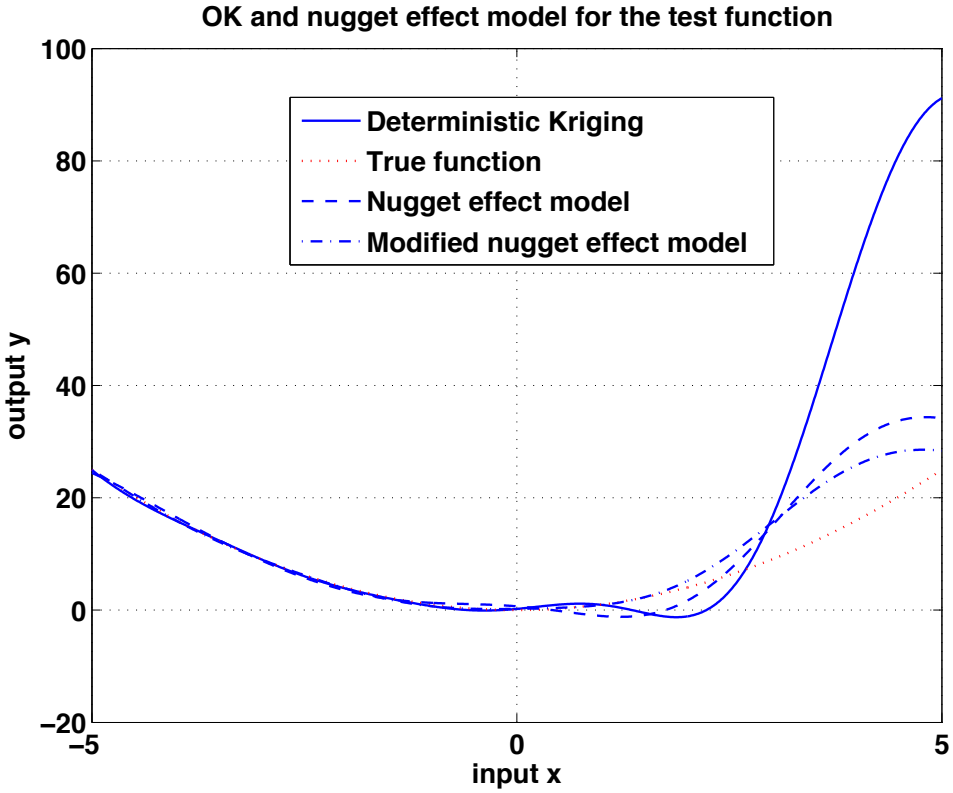


Figure 3.8: Different predictors' output for test function (r_{var} =1000).

The Average Absolute Error (AAE_S) is

$$AAE_S = \frac{1}{n+k}\sum_{i=1}^{n+k}|\hat{Z}(x_i) - Z(x_i)| \tag{3.34}$$

and also assesses the overall performance, and the Maximum Absolute Error (MAE_S)

$$MAE_S = \max|\hat{Z}(x_i) - Z(x_i)|_{i=1,2,\ldots,n+k} \tag{3.35}$$

reflects the presence of the poor prediction in local areas.

From this table, we note that the modified nugget-effect model is better than the nugget-effect model and the traditional ordinary kriging model. For all the models listed in Table 3.2, the error increases as the r_{var} increases; the second-order polynomial regression can greatly reduce the error by providing a better estimator for the process mean. Paired t-test results suggest that the differences

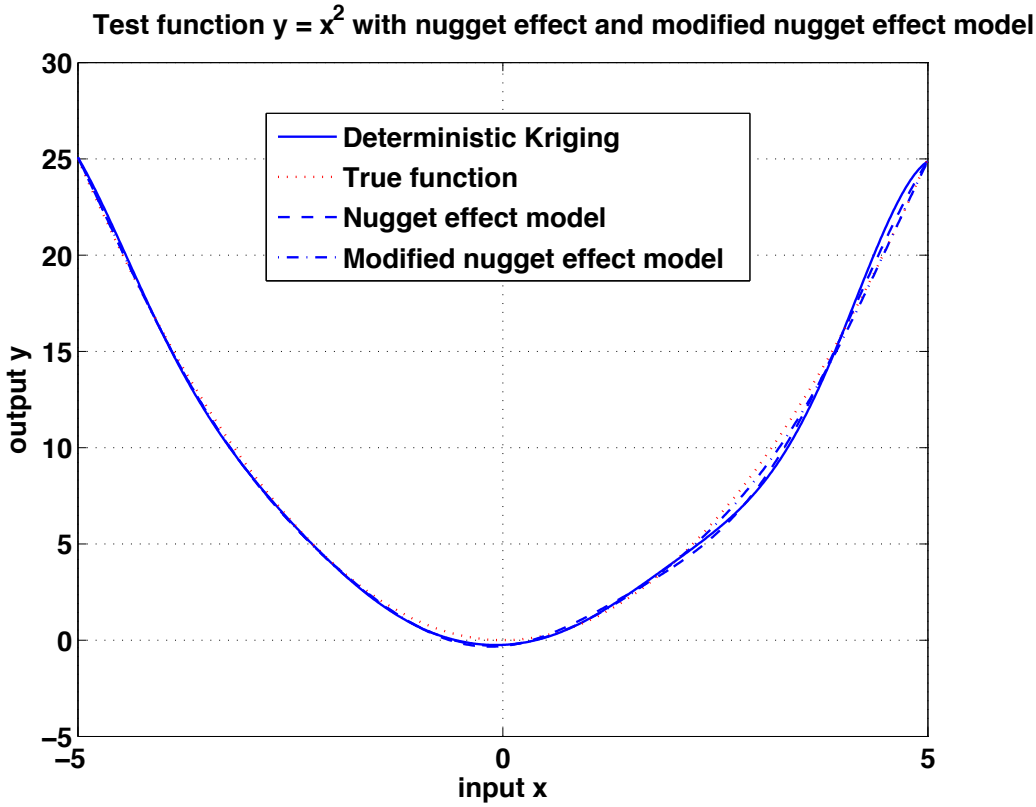


Figure 3.9: Different predictors' output for test function (r_{var}=100, 2nd-order polynomial regression model).

between the ordinary kriging model and the modified nugget-effect model, the differences between the nugget-effect model (max variance) and the modified nugget-effect model for all the measurements are statistically significant at alpha level of 0.05. The difference between the ordinary kriging model and the nugget-effect model (max variance) is statistically significant only for the high variance scenario (r_{var} = 100, 1000). As mentioned in section 3.3.1, for most cases, we tend to select a higher nugget-effect value in order to have a better estimator of ϕ_Z. For the nugget-effect model, ι_0 is set at the largest variance observed. To understand this selection, we study the impact of ι_0 on the prediction of the nugget-effect model. Figure 3.10 plots the influence of the nugget-effect value ι_0 on the estimation error. As can be seen, the larger nugget-effect value, the lower the MSE. This empirically verifies our observation in section 3.3.1.

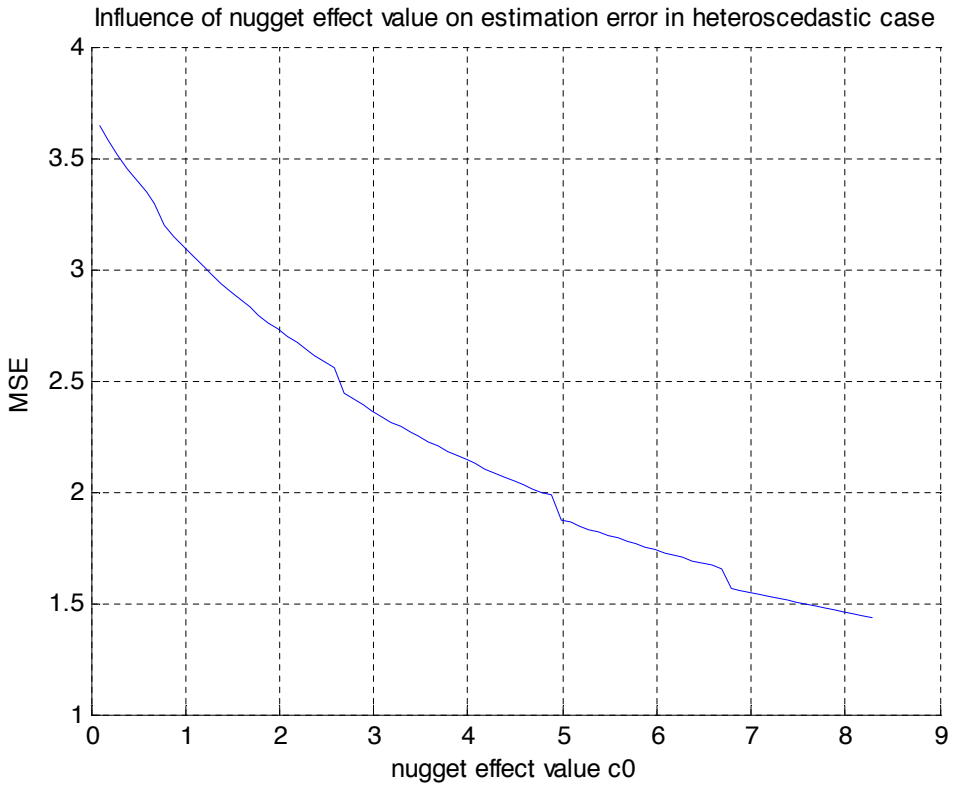


Figure 3.10: Influence of nugget value on MSE (test function).

3.4.2 M/M/1 queueing system

The M/M/1 queue system is a typical stochastic system which is widely studied in the literature. Because the expected waiting time for the M/M/1 system has a closed form, see Hillier & Lieberman (2001), it is easy to compare the model's performance with the true performance.

The data for this experiment is based on the M/M/1 queueing simulation system, which displays the heterogeneous variance characteristics. For the simulation system design, the input is the traffic load of the queueing system, and output is the expected waiting time over 10,000 customer arrivals with a warm up period of 2,000 customer arrivals. The input-output combinations of the M/M/1 system are used to build the model, with input points located at $x = 0.01, 0.05, 0.10, \ldots, 0.80, 0.85, 0.90$, totaling 19 points. To show the metamodel's performance, 1000 macro-replications are taken. For each replication, one krig-

ing metamodel is built based on the $n = 19$ observed input-output combinations. The output of the corresponding kriging predictor is generated at $k = 40$ evenly distributed points, $x = 0.02, 0.0425, 0.065, \ldots, 0.875, 0.8975$ (no extrapolations). Both the nugget-effect model and the modified nugget-effect model are tested and the comparisons of these two models are given in Table 3.3. Paired t-test results suggest that the differences between the nugget-effect model and the modified nugget-effect model are statistically significant at the alpha level of 0.05. This indicate that the modified nugget-effect model outperforms the nugget-effect model in all three measures for this M/M/1 queue example. The detrended studentization method proposed in Kleijnen & Beers (2005) takes replications at each location to obtain an average value, and standardizes it as the input. In order to make a comparison with the studentization method, we use the expected average waiting time instead of the average waiting time as the output of interest. We repeat this for 100 sub-groups, where in each sub-group, we obtain a sample size of 10 replications at each of the input points and calculate the sample means and the sample variances. For the studentization method, the sample mean and the sample variance are used to standardize the input. For the modified nugget-effect, both the sample mean and the sample variance are used as the inputs. Figure 3.11 and Figure 3.12 illustrate the studentization method and the modified nugget-effect model's performance on 100 sub-groups with the sample size of 10 per sub-group:

From Figure 3.11 and Figure 3.12, we see that the modified nugget-effect predictor's outputs are closer to the signal function than the studentization method. The performance measures MSES, AAES and MAES for the two methods are summarized in Table 3.4 for a variety of sample sizes. From Table 3.4, paired t-test results suggest that the differences between the studentization method and the modified nugget-effect model are statistically significant at the alpha level of 0.05 for sample size equals to 10 and 100. The modified nugget-effect model performs better in terms of the three measures. This is because the variability in the predictor is higher in the studentization method from the standardization and transformation of the predictor by its estimated signal function and variance. We note that the differences between the two methods decrease as the sample size for each subgroup increases. The improvement of the studentization method

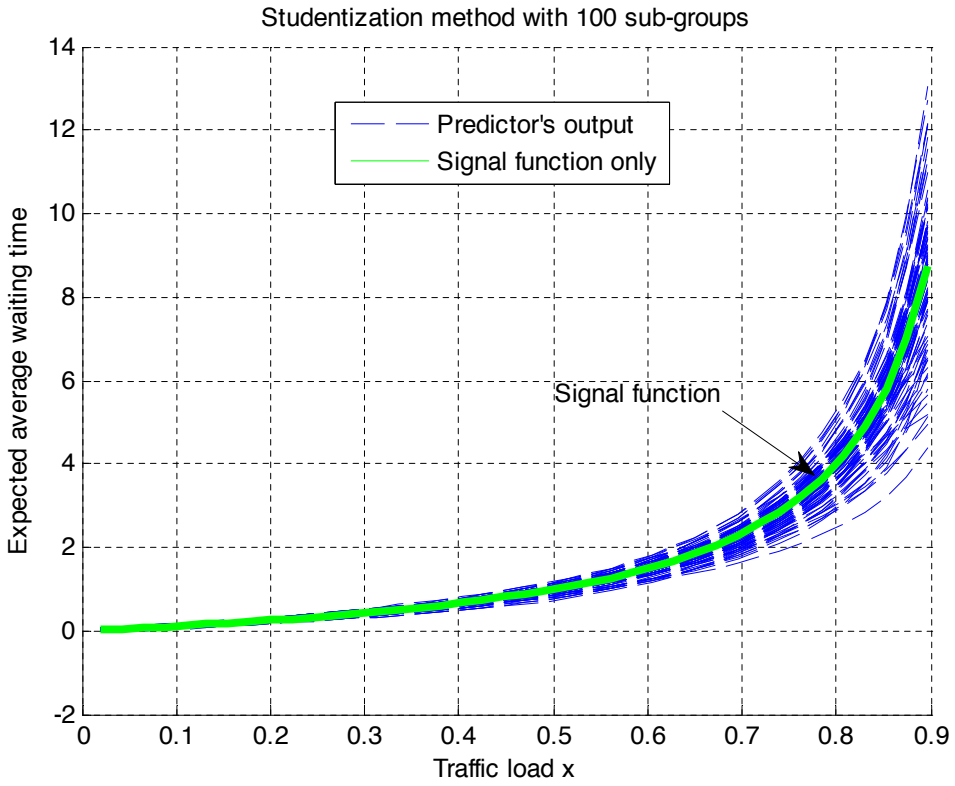


Figure 3.11: Studentization method with 100 sub-groups (sample size per sub-group = 10.

as the sample size increases and estimators improve is also noted in Kleijnen & Beers (2005). Hence, the modified nugget-effect model is a better choice when the total samples available are limited.

3.4.3 PAD system

In this example, we adopt the complex queueing network model studied in Cheng & Kleijnen (1999b). This packet assembly/disassembly device (PAD) is illustrated in Figure 3.13 below: The PAD system receives characters from several terminals. All the character arriving rates of the terminals are the same, and we denote this as x. The number of terminals is $N = 10$. The buffer size of each terminal is set at 32. Once the buffer is full, the characters in the buffer will be packed into a packet and sent to the output queue. If a special character is sent

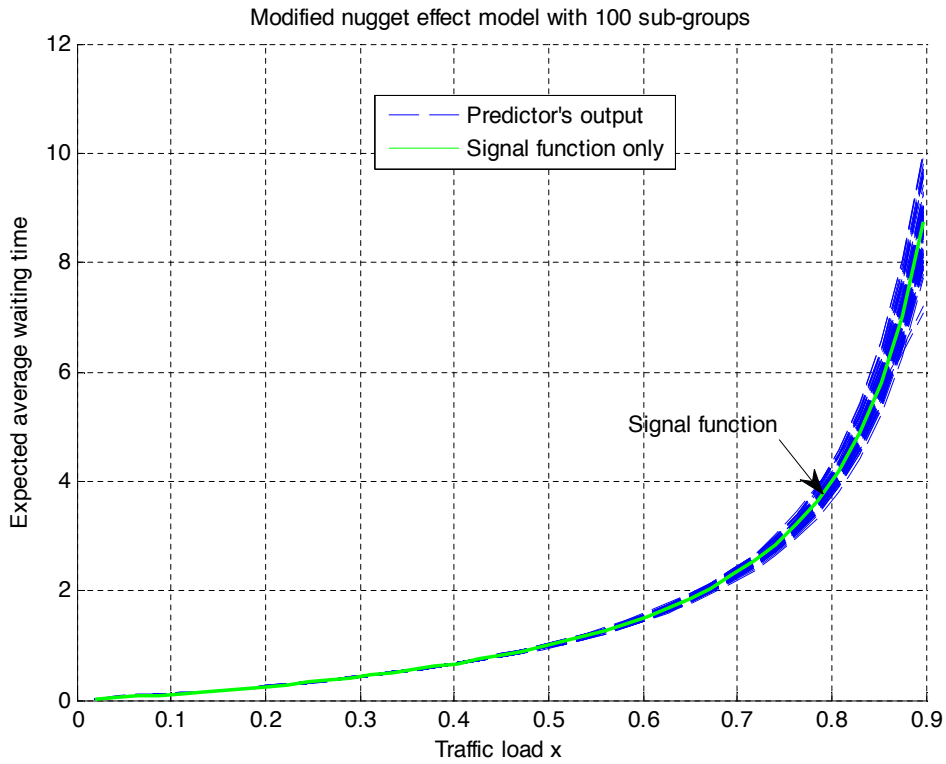


Figure 3.12: Modified nugget effect model with 100 sub-groups (sample size per sub-group = 10).

to the buffer, then all the characters in the buffer will be packed and sent to the output queue. The probability of a special character's arrival is 0.02. The packets at the output queue will be served in a FIFO rule and sent to the network at the speed of C char/sec. In this study, the input variable is the character arrival rate x while the output of interest is the average character delay T. This is the time interval from the moment the character arrives at the terminal to the moment the character leaves the output queue. Obviously, when the arrival rate x is low, the delay T will be high due to the long inter-arrival interval. In this study, we are interested in the higher saturation regions, so the focus region will be in the region $[0.5, 0.9]$. 9 observed points are evenly distributed in this region and 100 replications are taken for each observation location. Based on the estimated mean and variance of the 100 replications, we plot the prediction intervals below:

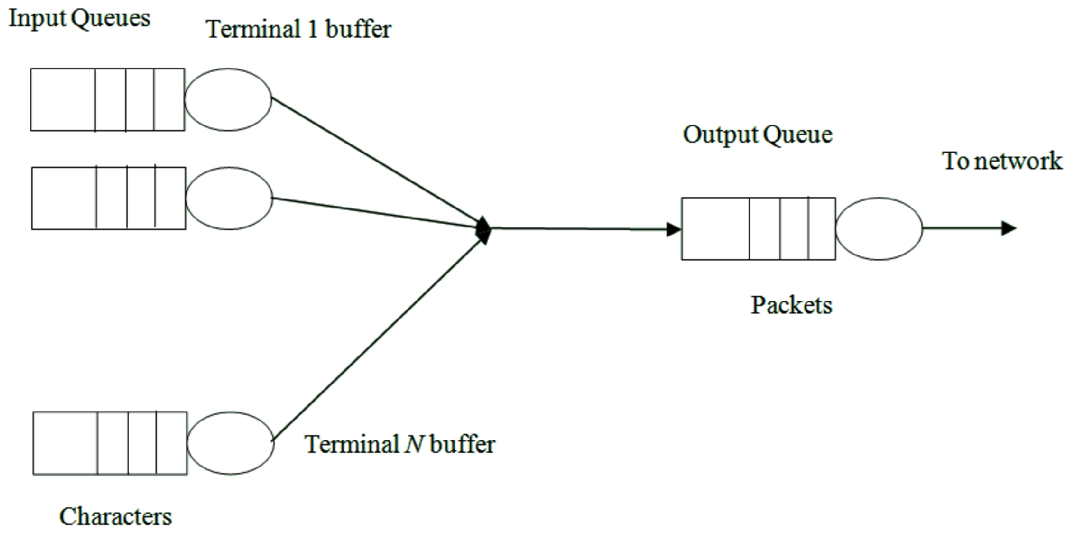


Figure 3.13: Queueing model for computer PAD system.

In Table 3.5, we use 41 evenly distributed unobserved test points to determine the performance of both the nugget-effect and the modified nugget-effect models. Paired t-test results show that the differences between nugget-effect model and the modified nugget-effect model are not statistically significant at the alpha level of 0.05. As the closed-form signal function is not known for this PAD system, the three measurements evaluate the differences between the single replication and the overall mean. From Figure 3.14, Figure 3.15 and Table 3.5, we see that the differences between the modified nugget-effect model and the nugget-effect model are not obvious. Although the modified nugget-effect model has a tighter prediction interval and is better able to reduce the influence of noise with high variance in the region between 0.85 and 0.9. Based on the test data collected at test points $x = 0.85, 0.86, 0.87, 0.88, 0.89, 0.90$. The differences between the modified nugget-effect model and the nugget-effect model are significant at alpha level of 0.05.

The numerical experiments in this section provide insights into the application of modified nugget-effect model in stochastic simulation. Although we focus on the development of our proposed model in this thesis, in practice it can be applied within sequential experimental design or optimization frameworks such as those proposed in Kleijnen (2009) and Huang *et al.* (2006). Because the proposed

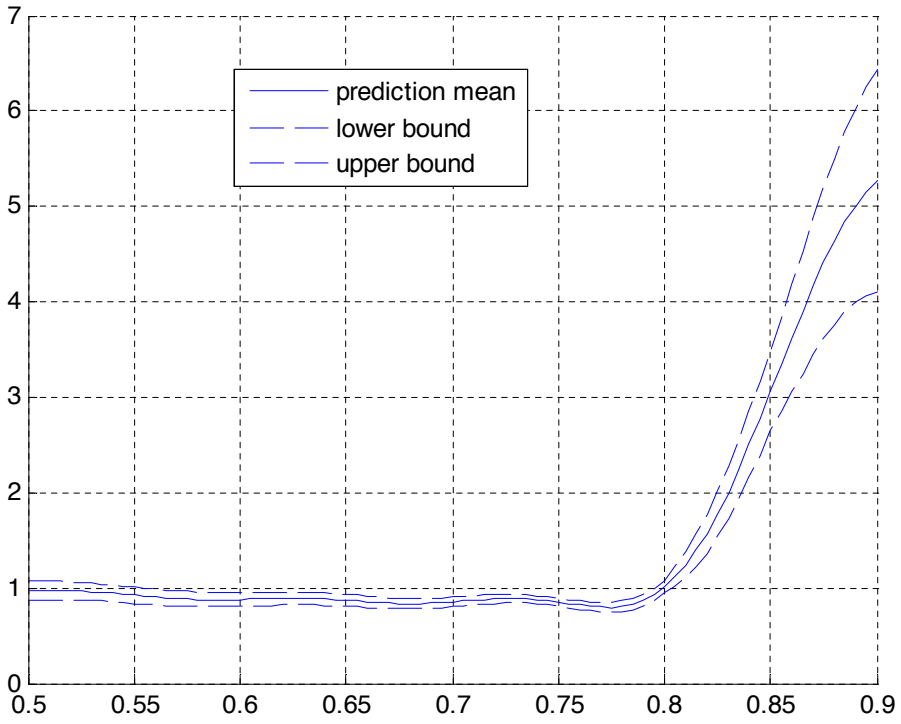


Figure 3.14: Prediction interval for nugget effect predictor (PAD system).

modified nugget-effect model requires prior information of the target simulation model, these stage-wise sequential approaches are useful where an initial stage can be used to estimate variances and structure, and subsequent follow up stages can be used to focus on refining the model in more interesting or promising regions. Hence the proposed model also can work in the real world scenarios with these stage-wise sequential approaches.

In conclusion, the results indicate that the modified nugget-effect model outperforms the stationary kriging model in heteroscedastic cases. It is a promising method for applications in simulation systems with heterogeneous variances. In the following chapter, we are going to further investigate the parameter estimation uncertainty in the following chapter.

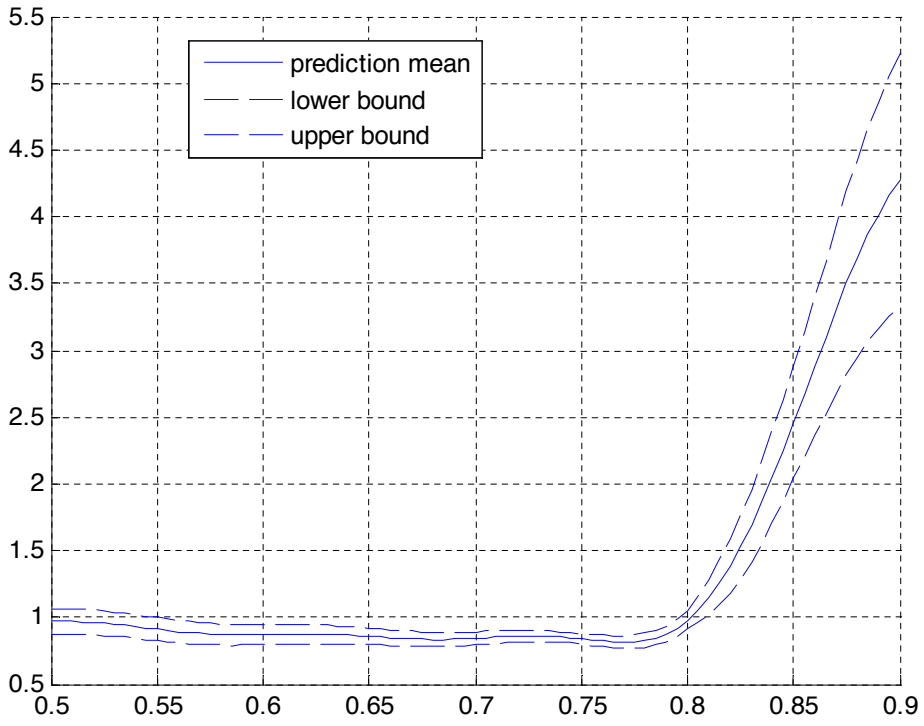


Figure 3.15: Prediction interval for modified nugget effect predictor (PAD system).

Table 3.2: Different error measures of different metamodels for the test function example.

Ratio(r_{var})	10			
Metamodels	MSE_{SO}	MSE_S	AAE_S	MAE_S
OK	0.990	2.572	0.780	3.315
Nugget-effect (min variance)	0.967	1.588	0.762	3.212
Nugget-effect (max variance)	0.927	1.426	0.664	2.758
Modified nugget-effect	0.646	0.971	0.513	1.992
Modified nugget-effect (optimal)	0.618	0.835	0.444	1.815
Ratio(r_{var})	100			
Metamodels	MSE_{SO}	MSE_S	AAE_S	MAE_S
OK	8.474	17.16	2.234	14.11
Nugget-effect (min variance)	7.836	11.11	1.836	9.743
Nugget-effect (max variance)	7.222	9.490	1.473	8.165
Modified nugget-effect	4.030	6.957	1.181	5.489
Modified nugget-effect (optimal)	3.340	5.725	0.930	5.153
Ratio(r_{var})	100(2nd poly)			
Metamodels	MSE_{SO}	MSE_S	AAE_S	MAE_S
OK	8.462	5.492	1.954	4.963
Nugget-effect (min variance)	7.402	0.513	0.738	2.854
Nugget-effect (max variance)	6.926	0.506	0.539	2.201
Modified nugget-effect	3.698	0.133	0.351	1.201
Modified nugget-effect (optimal)	2.902	0.092	0.198	1.065
Ratio(r_{var})	1000			
Metamodels	MSE_{SO}	MSE_S	AAE_S	MAE_S
OK	83.04	241.9	10.10	29.49
Nugget-effect (min variance)	66.13	108.3	9.469	14.40
Nugget-effect (max variance)	62.47	91.16	8.653	12.85
Modified nugget-effect	16.44	21.78	7.394	9.565
Modified nugget-effect (optimal)	12.39	17.15	5.535	7.921

Table 3.3: Error measurements of the nugget-effect model and the modified nugget-effect model for M/M/1 example.

	Nugget-effect model		Modified nugget-effect model	
	Mean	Var	Mean	Var
MSE_S	11.72	49.82	9.040	32.38
AAE_S	0.850	0.210	0.690	0.120
MAE_S	8.560	32.83	6.080	18.35

Table 3.4: Error measurements of the nugget-effect model and the modified nugget-effect model for M/M/1 example.

Sample size per sub-group	Studentization method		Modified nugget-effect model	
		Mean	Var	Mean	Var
10	MSE_S	2.2040	2.6041	0.6525	0.2162
	AAE_S	0.1795	0.0149	0.0454	0.0009
	MAE_S	1.4373	1.2573	0.4584	0.114
100	MSE_S	1.5485	0.9168	0.4677	0.1110
	AAE_S	0.1695	0.0103	0.0223	0.00007
	MAE_S	0.7417	0.2646	0.4299	0.1047
10	MSE_S	0.5444	0.1167	0.3966	0.0406
	AAE_S	0.0568	0.0013	0.0216	0.00002
	MAE_S	0.2793	0.0343	0.3377	0.0390

Table 3.5: Error measurement of the nugget-effect model and the modified nugget-effect model for the PAD system example.

	Nugget-effect model		Modified nugget-effect model	
	Mean	Var	Mean	Var
MSE_S	0.72	3.26	0.65	2.93
AAE_S	0.82	0.15	0.76	0.11
MAE_S	1.21	2.81	0.98	2.56

Chapter 4

PARAMETER ESTIMATION FOR KRIGING METAMODEL IN STOCHASTIC SIMULATION

4.1 Introduction

As illustrated in Section 3.2.3, the parameter estimation uncertainties can be affected by the random noise in the stochastic simulation scenario. In an ideal situation, these parameters are assumed known. In practice however, these parameters can only be estimated from sample data, making them random variables dependent on the experimental design and sample observations. Moreover, the prediction error of the kriging model, which is a commonly used quality measure of the fit and accuracy of the model, is a function of sample data and model parameters. In the ideal case when the parameters are known, the prediction error measures the "true" prediction error. Typically however, the parameters are unknown and the prediction error is estimated by replacing the unknown parameters with point estimates. This "plug-in" estimator will however underestimate the true prediction error as it does not take into account the uncertainty of the model parameters. In some cases, it can cause overconfidence in the predictors. This additional error is also noted in Cressie (1993).

Similar problems in parameter uncertainty have been studied in time series models, heteroscedastic regression models, mixed linear models (Khatri & Shah (1981), Kackar & Harville (1984), Reinsel (1984)) and general linear models (Toyooka (1976), Harville (1985) and Harville & Jeske (1992)). Zimmerman & Cressie (1992a) extended this research in parameter estimation uncertainty to the mean squared prediction error (MSPE) of spatial linear models, and examined the appropriateness of the "plug-in" estimator of the MSPE with alternative approximations. den Hertog *et al.* (2005) studied the similar problem with the sensitivity parameter ϕ of the deterministic kriging model with a boot-strapping approach, and showed that the traditional kriging variance underestimates the true kriging variance as noted in Cressie (1993). These studies demonstrate that the problem of parameter uncertainty in spatially correlated models can have a large influence on the prediction error in the deterministic (noiseless) case. In stochastic simulations, this problem can be amplified by the random noise $\varepsilon(x)$ of the system as it increases the variability of the parameter estimates (see illustrative example below). In this chapter, we extend previous studies to look at the effects of stochastic noise on parameter estimation and the overall prediction error. We further decompose the mean squared error into individual components reflecting parameter uncertainty, response model misspecification and the stochastic noise. To illustrate the influence of random noise in stochastic simulation on parameter estimation and prediction error, consider the example where the response is $Y(x) = \sin(x) + \varepsilon(x)$, and $\varepsilon \sim N(0, \sigma_\varepsilon^2(x))$. Observations are obtained at seven equally spaced input locations from $[0, 2\pi]$ for various levels of $\sigma_\varepsilon^2(x)$. Suppose a deterministic kriging model with an exponential correlation function is used to fit the data. For each level of $\sigma_\varepsilon^2(x)$, 1000 replications of observations are taken. For each replication, we estimate the sensitivity parameter ϕ_Z of the exponential correlation function and compute the overall prediction error at the point $x_0 = 3\pi/4$, where the overall prediction error is the squared differences between the estimated kriging predictor and the mean function $\mathrm{E}[Y(x_0)] = S(x_0) = \sin(x_0)$. Table 4.1 summarizes the results. From Table 4.1, we see that both of the variance of the estimated ϕ and averaged overall prediction error increase as the variance of the random noise $\varepsilon(x)$ increases. As the input design for x is fixed for all four noise levels, the inherent model misspecification error is the same throughout,

Table 4.1: Numerical results on the parameter estimation and prediction error.

$\sigma_\varepsilon^2(x)$	Variance of the ϕ_Z estimator	Averaged Overall Prediction Error
0.1	9.161E-12	0.0159
0.2	5.044E-11	0.0268
0.5	1.181E-8	0.1335
1.0	1.894E-7	0.5923

and hence, the increase in the overall prediction error can be attributed to the noisy sample data. The results from this example show that the performance of deterministic kriging model worsens as the noise level increases. As stated in Chapter 3, the kriging model with nugget effect (Nugget effect model) has been proposed for the application with homoscedastic case, and the kriging model with modified nugget effect (Modified nugget effect model) and the stochastic kriging model are proposed to be applied in the heteroscedastic case. In this chapter, we will look more closely at the influence of random noise in stochastic simulations on these three different model types: traditional deterministic kriging mode, Cressie (1993); nugget effect kriging model, Cressie (1993); and the modified nugget effect kriging model, Yin *et al.* (2008). We first decompose the prediction error into components reflecting the model misspecification error, parameter estimation error and the stochastic error. We will then examine a simple two-point tractable problem and provide some insights on the effects of the random noise in stochastic simulations on the individual components of the three different model types. Finally, we provide a numerical study on three additional examples.

4.2 Decomposition of the Overall Prediction Error for Stochastic Case

The additional error caused by parameter estimation uncertainty is reported in several different research papers. den Hertog *et al.* (2005) discussed the parameter

estimation uncertainty issues for kriging metamodel, especially on the parameter estimation uncertainty's influence on the overall prediction error. Bootstrapping numerical experiments show that the actual kriging model prediction error which correctly accounts for parameter estimation uncertainty is larger than the traditional kriging variance given the known parameter. Assuming the predictor $\hat{Z}(x_0)_Y$ is a linear function of the parameter ϕ_Z and the estimator for ϕ_Z is unbiased, from the results of Kackar & Harville (1984), the prediction error with unknown parameter can be approximated by Equation (4.1).

$$MSE[\hat{Z}(x_0)_Y] = tr[A(\hat{\phi}_Z)B(\hat{\phi}_Z)] + MSE[\hat{Z}(x_0)_Y|\hat{\phi}_Z] \tag{4.1}$$

where $tr[A(\hat{\phi}_Z)B(\hat{\phi}_Z)]$ is an approximation of the additional error introduced when the estimator $\hat{\phi}_Z$ is used, $A(\hat{\phi}_Z) = var\left(\frac{\partial \hat{Z}(x_0)_Y}{\partial \phi_Z}\right)$, and$B(\hat{\phi}) = \mathrm{E}[(\hat{\phi}_Z - \phi_Z)(\hat{\phi}_Z - \phi_Z)^{\mathrm{T}}]$. The second component of the right-hand side of Equation (4.1) is the traditional mean squared error (MSE) when $\hat{\phi}_Z$ is used, and this can be further decomposed as:

$$\begin{aligned} MSE[\hat{Z}(x_0)_Y] &= \mathrm{E}\left[(\sum_{k=1}^{m}\lambda_k Y_k - Z(x_0))^2|\hat{\phi}_Z\right] \\ &= \mathrm{E}\left[(\sum_{k=1}^{m}\lambda_k(Z(x_k)+\varepsilon(x_k)) - Z(x_0))^2|\hat{\phi}_Z\right] \\ &= \mathrm{E}\left[(\sum_{k=1}^{m}\lambda_k Z(x_k) - Z(x_0) + \sum_{k=1}^{m}\lambda_k\varepsilon(x_k))^2|\hat{\phi}_Z\right] \\ &= \mathrm{E}\left[(\sum_{k=1}^{m}\lambda_k Z(x_k) - Z(x_0))^2|\hat{\phi}_Z\right] + \sum_{k=1}^{m}\mathrm{E}[\lambda_k^2|\hat{\phi}_Z]\sigma_\varepsilon^2(x_k) \end{aligned} \tag{4.2}$$

The first term on the right hand side of Equation (4.2) is the prediction error caused by model misspecification. The second term is the direct effect of the stochastic noise $\varepsilon(x)$ on the prediction error, and is a function of the variance of $\varepsilon(x)$. Combining Equation (4.1) and Equation (4.2) together, we find that the overall prediction error can be decomposed into the following three error

components:

$$MSE[\hat{Z}\,(x_0)_Y] = \mathrm{E}\left[\left(\sum_{k=1}^{m}\lambda_k Z(x_k) - Z(x_0)\right)^2 |\hat{\phi}_Z\right] + \sum_{k=1}^{m}\mathrm{E}[\lambda_k^2|\hat{\phi}_Z]\sigma_\varepsilon^2(x_k) + tr[A(\hat{\phi}_Z)B(\hat{\phi}_Z)] \tag{4.3}$$

The prediction error caused by model misspecification is inherent in the meta-model selection and will not be the focus in this research. In the next section, we analyze how the random noise $\varepsilon\,(x)$ affects the parameter estimation of ϕ_Z and the additional error caused by parameter estimation uncertainty (the last term in Equation (4.3)).

4.3 Maximum Likelihood Estimation with Stochastic Response

The Maximum Likelihood Estimation (MLE) method is commonly used in kriging model's estimation. Assuming a Gaussian random process, the log-likelihood function is given by:

$$\ell(\mu, \phi_Z, \sigma_Z^2) = -\frac{1}{2}\ln\det(R') - \frac{1}{2}(\bar{Y} - F\beta)^{\mathrm{T}} R'^{-1}(\bar{Y} - F\beta) \tag{4.4}$$

where F is the design matrix for the ordinary least squares model, β is the regression parameters and $F\beta$ represents the mean function μ; σ_Z^2 is the variance of the Gaussian random process Z, which indicates the variability of an unknown point in Z. Here we write the correlation matrix as $R(\phi_Z)$ to denote that it is a function of parameter ϕ_Z. We can find the estimators for μ, ϕ_Z and σ_Z^2 by taking the first order derivatives:

$$\frac{\partial\ell(\mu, \phi_Z, \sigma_Z^2)}{\partial\mu} = 0, \frac{\partial\ell(\mu, \phi_Z, \sigma_Z^2)}{\partial\phi_Z} = 0, \frac{\partial\ell(\mu, \phi_Z, \sigma_Z^2)}{\partial\sigma_Z^2} = 0$$

Solving the above three equations, the MLE estimators μ and σ_Z^2 result as functions of ϕ. For simplification purposes, μ and σ_Z^2 are typically assumed fixed or known in order to estimate the sensitivity parameter ϕ_Z. This simplifies the likelihood function to a function of ϕ_Z only. However, in this simplification, the

MLE estimator for ϕ_Z is biased as the estimation of ϕ_Z depends also on β, which is usually unknown, see Cressie (1993). In order to simplify the approximation in Equation (4.1), we use instead the restricted maximum likelihood (REML) proposed by Patterson & Thompson (1971) and Patterson & Thompson (1974) which provides an unbiased estimator of ϕ_Z for this Gaussian random process. With this unbiased estimator, $B(\phi_Z)$ will equal to the variance of the ϕ_Z estimator. The log-likelihood function for the REML is given by:

$$\begin{aligned} \ell(\phi_Z, \sigma_Z^2) &= \frac{1}{2}|F^{\mathrm{T}}F| - \frac{1}{2}\log|\sigma_Z^2 R(\phi_Z)| - \frac{1}{2}\log|F^{\mathrm{T}}(\sigma_Z^2 R(\phi_Z))^{-1}F| \\ &- \frac{1}{2}Y^{\mathrm{T}}((\sigma_Z^2 R(\phi))^{-1} - \frac{(\sigma_Z^2 R(\phi))^{-1}FF^{\mathrm{T}}(\sigma_Z^2 R(\phi))^{-1}}{F^{\mathrm{T}}(\sigma_Z^2 R(\phi))^{-1}F})Y \end{aligned} \tag{4.5}$$

which is independent of β. The difference between Equation (4.4) and Equation (4.5) is especially significant in the small sample case, like the two-point problem we propose here, see Cressie (1993).

4.3.1 A simple two-point problem

We design a simple two-point problem to provide some theoretical insights to the parameter estimation problem for the kriging model: Points P_0, P_1 and P_2 are

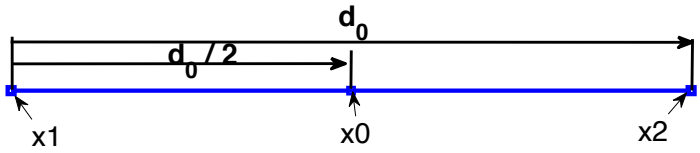


Figure 4.1: Design for two-point problem.

evenly spaced. Take points P_1 and P_2 to be the observation points, and point P_0 to be the prediction point. Suppose that the mean function Z is an unknown Gaussian random process with the mean function $\mu(x)$ and bias function $\delta(x)$.

The additional random noise function $\varepsilon(x)$ follows the normal distribution with zero mean and unknown heterogeneous variance function $\varepsilon(P_i)$. Then at points:

$$P_1(x_1) : Y_1 = Z_1 + \varepsilon_1, \quad P_1(x_1) : Y_2 = Z_2 + \varepsilon_2 \tag{4.6}$$

In the next subsection, we describe the parameter estimation techniques for the unknown parameter before discussing its effects on the overall prediction error.

For this two-point problem, the terms in Equation (4.5) are given as:

$$\sigma_Z^2 = 1, \quad F = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad R(\phi) = \begin{pmatrix} 1 & \exp(-d\phi_Z) \\ \exp(-d\phi_Z) & 1 \end{pmatrix}$$

As mentioned in Section 4.2.1, the kriging predictor is not a linear function of the parameter ϕ in this two-point case. In order to make the predictor a linear function of the estimated parameter to apply the approximation in Equation (4.1), a reparameterization is made as follows:

$$\rho = \exp(-d\phi_Z), \quad R(\rho) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

With this reparameterization, Equation (4.5) becomes a function of ρ instead of ϕ:

$$\begin{aligned} L_w(\rho) &= \frac{1}{2}|F^{\mathrm{T}}F| - \frac{1}{2}\log|R(\rho)| - \frac{1}{2}\log|F^{\mathrm{T}}R(\rho)^{-1}F| \\ &- \frac{1}{2}Y^{\mathrm{T}}\left[R(\rho)^{-1} - \frac{R(\rho)^{-1}FF^{\mathrm{T}}R(\rho)^{-1}}{F^{\mathrm{T}}R(\rho)^{-1}F}\right]Y \end{aligned} \tag{4.7}$$

4.3.2 Analytical Results

For the three different model forms described in Section 4.1, we have the following results: Deterministic kriging model:

$$\hat{\rho} = 1 - \frac{(y_1 - y_2)^2}{2}$$

Nugget effect model:

$$\hat{\rho}_N = \frac{2 + 2\iota_0 - 2(y_1 - y_2)^2}{2}$$

Modified nugget effect model:

$$\hat{\rho}_M = \frac{2 + \iota_1^* + \iota_2^* - 2(y_1 - y_2)^2}{2}$$

Following, the expectation and variance of the parameter estimators are given as:

$$\mathrm{E}(\hat{\rho}) = 1 - \frac{1}{2}\mathrm{E}((y_1 - y_2)^2) \tag{4.8}$$

$$var(\hat{\rho}) = \frac{1}{4}var((y_1 - y_2)^2) \tag{4.9}$$

From Equation (4.6), it is straightforward to see that y_1 follows normal distribution with mean z_1 and variance $\sigma_\varepsilon^2(x_1)$, and y_2 follows normal distribution with mean z_2 and variance $\sigma_\varepsilon^2(x_2)$. Separating the mean and pure noise components, we get:

$$\mathrm{E}((y_1 - y_2)^2) = \sigma_\varepsilon^2(x_1) + \sigma_\varepsilon^2(x_2) + (z_1 - z_2)^2 \tag{4.10}$$

$$\begin{aligned} var((y_1 - y_2)^2) &= 2(\sigma_\varepsilon^2(x_1) + \sigma_\varepsilon^2(x_2))^2 + 4(\sigma_\varepsilon^2(x_1) \\ &+ \sigma_\varepsilon^2(x_2))(z_1 - z_2)^2 \end{aligned} \tag{4.11}$$

Combining Equation (4.8) and Equation (4.10),

$$\mathrm{E}((y_1 - y_2)^2) = 1 - \frac{1}{2}(\sigma_\varepsilon^2(x_1) + \sigma_\varepsilon^2(x_2) + (z_1 - z_2)^2) \tag{4.12}$$

$$\begin{aligned} var((y_1 - y_2)^2) &= \frac{1}{2}(\sigma_\varepsilon^2(x_1) + \frac{1}{4}(x_2))^2 + 4(\sigma_\varepsilon^2(x_1) \\ &+ \sigma_\varepsilon^2(x_2))(z_1 - z_2)^2 \end{aligned} \tag{4.13}$$

Similarly, for the nugget effect model and modified nugget effect model, we obtain the following:

$$\mathrm{E}((y_1 - y_2)^2) = 1 + \iota_0 - \frac{1}{2}(\sigma_\varepsilon^2(x_1) + \sigma_\varepsilon^2(x_2) + (z_1 - z_2)^2) \tag{4.14}$$

$$\begin{aligned} var((y_1 - y_2)^2) &= \frac{1}{2}(\sigma_\varepsilon^2(x_1) + \frac{1}{4}(x_2))^2 + 4(\sigma_\varepsilon^2(x_1) \\ &+ \sigma_\varepsilon^2(x_2))(z_1 - z_2)^2 \end{aligned} \tag{4.15}$$

$$\mathrm{E}((y_1 - y_2)^2) = 1 + \frac{\iota_1}{2} + \frac{\iota_2}{2} - \frac{1}{2}(\sigma_\varepsilon^2(x_1) + \sigma_\varepsilon^2(x_2) + (z_1 - z_2)^2) \tag{4.16}$$

$$\begin{aligned} var((y_1 - y_2)^2) &= \frac{1}{2}(\sigma_\varepsilon^2(x_1) + \frac{1}{4}(x_2))^2 + 4(\sigma_\varepsilon^2(x_1) \\ &+ \sigma_\varepsilon^2(x_2))(z_1 - z_2)^2 \end{aligned} \tag{4.17}$$

For deterministic model, we see from Equation (4.12) that the expectation and variance of the estimated parameter are functions of the input variance $\sigma_\varepsilon^2(x_1)$

and $\sigma_\varepsilon^2(x_2)$. If the input variances increase, the variance of the estimated parameter will also increase and its mean will decrease, indicating a weaker correlation between the points. From Equation (4.12), the expectation of the estimated parameter can be negative if the variance $\sigma_\varepsilon^2(x_1)$ and $\sigma_\varepsilon^2(x_2)$ are high enough. However as we assume the exponential correlation function in this two-point problem, we consider only non-negative correlations. For the cases when estimated ρ is negative, the restricted likelihood function is monotonically decreasing, indicating that extra sample data is needed. For the modified nugget effect model, from Equation (4.16), we see that the influence of the input variance can cancel out if ι_1^* and ι_2^* are the exact estimators of $\sigma_\varepsilon^2(x_1)$ and $\sigma_\varepsilon^2(x_2)$. Similarly, from Equation (4.14), we see that the nugget effect model can have the same results when the nugget value c0 equals to the average of $\sigma_\varepsilon^2(x_1)$ and $\sigma_\varepsilon^2(x_2)$. This partially explains what was observed in Yin *et al.* (2008), where it was observed that the estimated ϕs for the modified nugget effect model and nugget effect model is closer to the optimal value than the deterministic model.

4.3.3 Influence of Parameter Estimation on Overall Prediction Error

From the approximation in Equation (4.1), the additional prediction error caused by parameter estimation uncertainty can be approximated as $tr[A(\hat{\phi}_Z)B(\hat{\phi}_Z)]$, where $tr[Q]$ stands for trace of matrix Q. Based on the REML, can be computed and is the variance of $\hat{\rho}$, see Harville (1985). As a result, we can formulate the approximation as a function of ρ:

$$tr[A(\hat{\phi}_Z)B(\hat{\phi}_Z)] = \frac{1-\hat{\rho}}{2} var(\hat{\rho})$$

Since the variance of the estimator is the same for all three models, an estimator $\hat{\rho}$ closer to 1 is favorable. Comparing Equation (4.12) and Equation (4.16), the modified nugget effect estimator $\hat{\rho}_M$ is closer to one in expectation than the estimated parameter given with deterministic kriging model. With careful selection of the nugget value ι_0 in Equation (4.14), the additional error incurred by the nugget effect model can be as small as the modified nugget effect model. In this simple example, we see that in stochastic situations where random noise is

present, selection of the appropriate stochastic model can reduce the additional error introduced by parameter estimation in ϕ_Z. Furthermore, although not addressed in this thesis, good knowledge or accurate estimation of $\sigma_\varepsilon^2(z)$ can also improve the estimation error.

4.4 Numerical Experiments

The two-point problem given in the Section 4.4 illustrates how the random noise $\varepsilon(x)$ affects the parameter estimation uncertainty and increases the additional prediction error in the end. In order to simultaneously study the effects of the random noise on the three individual error components, three numerical examples are further studied in this section. The first example is a one-dimension problem with a step variance function studied in Yin *et al.* (2008). In this example, the correlation is strong in the noiseless case. The second example extends to a two-dimension problem with a step variance function. Several different noise level scenarios are studied in this example. The third example is more complex functional form taken from Hussain *et al.* (2002b) with a continuous heterogeneous variance function. For each numerical example, 1000 replications are applied. The mean and variance of the ϕ_Z estimator, prediction error caused by model misspecification, additional prediction error caused by noisy data and additional prediction error caused by parameter estimation uncertainty are provided based on all the 1000 replications. For notation convenience, Table 4.2 summarizes the abbreviations used in the following numerical experiments. The ErrA, ErrM, ErrN and ErrP refer to the individual components of the prediction error decomposition in Equation (4.3). ϕ_Z^* is the ϕ_Z estimated based on noiseless observations. The comparison between ϕ_Z^* and $\hat{\phi}_Z$ estimated based on the noisy sample data can provide an insight of parameter estimation uncertainty in the stochastic case.

4.4.1 One Dimension Quadratic Test Function

The quadratic test function $Y = x^2 + \varepsilon(x)$ was used in Section 3. $\varepsilon(x)$ is the additional noise function with step variance $\sigma_\varepsilon^2(x) = 0.083$ when $x \in [-5, 2)$, $\sigma_\varepsilon^2(x) = 8.3$ when $x \in [2, 5]$. Observation points are located at

Table 4.2: Abbreviations for notations.

DK	Deterministic kriging model
NK	Nugget effect kriging model
MK	Modified nugget effect kriging model
ErrA	Overall prediction error
ErrM	Prediction error caused by model misspecification
ErrN	Additional prediction error caused by noisy data
ErrP	Additional prediction error caused by parameter estimation uncertainty (approximation)
ϕ_Z^*	The estimated ϕ_Z based on noiseless observations

$x = [-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5]$, and the prediction point is located at $x_0 = [-0.5]$. The results are given in Table 4.3. From Table 4.3, we see that the modified nugget effect model helps in reducing the overall prediction error ErrA. Both the prediction error caused by the noisy data and parameter estimation uncertainty for the modified nugget effect model are lower than the other two models. Considering the ErrP to ErrA ratio, an indicator of the fraction of the overall error attributed to the parameter estimation uncertainty, we see that the modified nugget effect model has a lower ratio than the other two models. As for the estimator, we see from the last two rows that the ϕ_Z^* and prediction error caused by model misspecification are very small. This indicates that there is sufficient samples in the sample space and the correlation is strong in the noiseless case. The estimated ϕ_Z is higher than the ϕ_Z^* for all three models, which indicates the correlation is weakened by the random noise. Comparing all the three estimated ϕ_Z's, the modified nugget effect model has the estimator the closest to ϕ_Z^*.

Table 4.3: Error measurements of the nugget-effect model and the modified nugget-effect model for M/M/1 example.

Estimation based on noisy data					
	$\hat{\phi}_Z$				
	Mean	Var	ErrA	ErrN	ErrP
DK	0.832	5.261	1.375	1.102	0.123
NK	0.245	0.625	0.573	0.427	0.003
MK	0.007	0.001	0.294	0.205	2E-7
	Mean	Var	ErrA	ErrN	ErrP
Estimation based on noiseless data					
	ϕ_Z^*			ErrM	
	0.002			5E-13	

4.4.2 Two Dimension Linear Function

The two dimension linear function is given as follows: $Z(x_1, x_2) = x_1 + x_2$ kriging model: As seen in Figure 4.2, observation points are located at $P_i(x_1, x_2) = [P_1(1,1), P_2(0,1), P_3(-1,1), P_4(-1,0), P_5(-1,-1), P_6(0,-1), P_7(1,-1), P_8(1,0)]$, and the prediction point is located at $P_0(x_1, x_2) = [(0,0)]$. To find out how the three models perform as the variance levels at the different locations increase, several different noise level scenarios for $\sigma_\varepsilon^2(x_i), i = 1, 2, \ldots, 8$ (shown in Table 4.4) are tested. The results are given in

Table 4.4: Noise level scenarios for two dimension linear test function.

Scenarios	$\sigma_\varepsilon^2(x_1)$	$\sigma_\varepsilon^2(x_2)$	$\sigma_\varepsilon^2(x_3)$	$\sigma_\varepsilon^2(x_4)$	$\sigma_\varepsilon^2(x_5)$	$\sigma_\varepsilon^2(x_6)$	$\sigma_\varepsilon^2(x_7)$	$\sigma_\varepsilon^2(x_8)$
L1	0.5	0.5	0.1	0.1	0.1	0.1	0.1	0.1
L2	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.1
L3	0.5	0.5	0.5	0.5	0.5	0.5	0.1	0.1

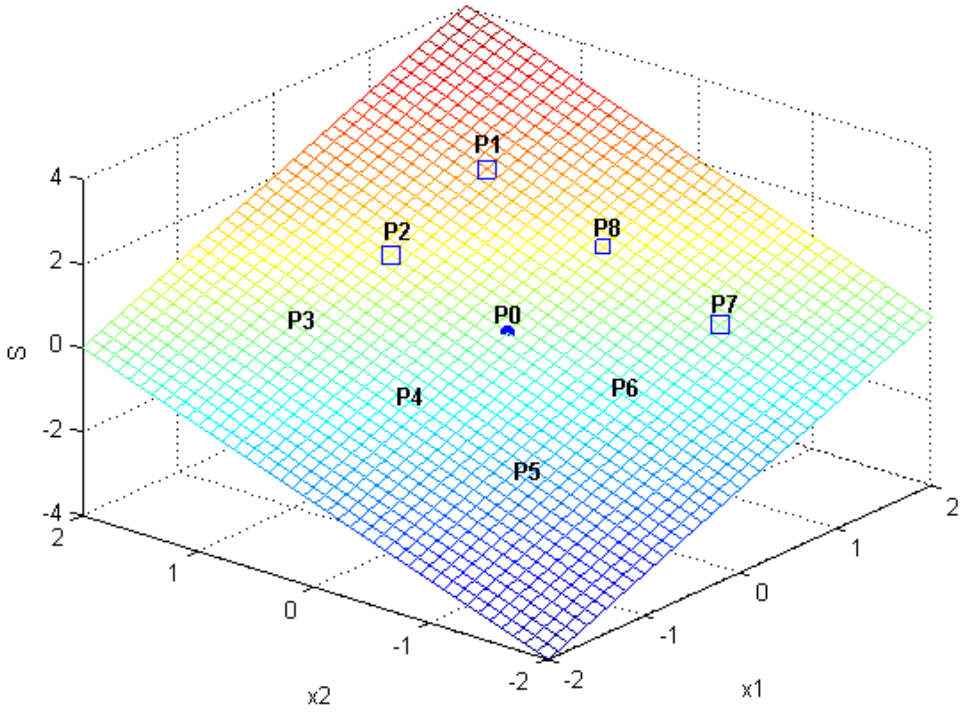


Figure 4.2: Two dimension linear test function.

Similar results are observed in Table 4.5. The modified nugget effect model performs better that the nugget effect model and deterministic kriging model in all the three scenarios. As the variance of the input random noise $\varepsilon(x)$ increases from L1 to L3, the overall prediction error and all the additional error components increase for all three models. The ErrP to ErrA ratio also increases from L1 to L3. In L1 where the variance of the random noise is relatively low, the ErrP is not significant considering the scale of ErrN. In L3 the ErrP cannot be ignored as it consists of about 20% of the ErrA.

4.4.3 Two Dimension Sinusoidal Function

The sinusoidal function is taken from Hussain *et al.* (2002b). The plot of the function and design are given in.

Table 4.5: Results for the two dimension linear test function.

Estimation based on noisy data						
		$\hat{\phi}_Z$				
		Mean	Var	ErrA	ErrN	ErrP
DK	L1	0.10	0.01	0.06	0.05	2E-4
	L2	0.18	0.02	0.10	0.09	0.01
	L3	0.36	10.0	0.21	0.13	0.05
NK	L1	0.09	4E-3	0.03	0.02	1E-4
	L2	0.10	0.01	0.06	0.05	3E-3
	L3	0.17	4.24	0.11	0.07	0.02
MK	L1	0.04	2E-3	0.01	0.01	6E-5
	L2	0.07	0.01	0.04	0.03	1E-3
	L3	0.17	3.40	0.08	0.06	0.02
Estimation based on noiseless data						
		ϕ_Z^*			ErrM	
		1E-3			1E-14	

The mathematical form of sinusoidal function is:

$$Y(x_1, x_2) = x_1 \sin(x_2) + x_2 \sin(x_1) + \varepsilon(x_1, x_2)$$

The additional noise component $\varepsilon(x)$ follows normal distribution with zero mean and variance function $\sigma_\varepsilon^2(x) = x_1^2$. Observation points are located at $P_i(x_1, x_2) = [P_1(0, 2), P_2(-1, 1), P_3(1, 1), P_4(-2, 0), P_5(2, 0), P_6(-1, -1), P_7(1, -1), P_8(0, -2)]$, prediction points are located at $P_j(x_1, x_2) = [P_9(0, 0), P_{10}(2, -1)]$. The results are given in the following Table 4.6. From the results in Table 4.6, we know that the correlation is rather weak as ϕ_x^* is large. For the interpolation at P_9, the overall prediction error is low compared with the extrapolation case at P_{10}. The prediction error caused by model misspecification in P10 is considerably larger than the one at P_9 due to the lack of information for extrapolation. In this case, considering the more appropriate ErrP to ErrN ratio, we see the additional prediction error

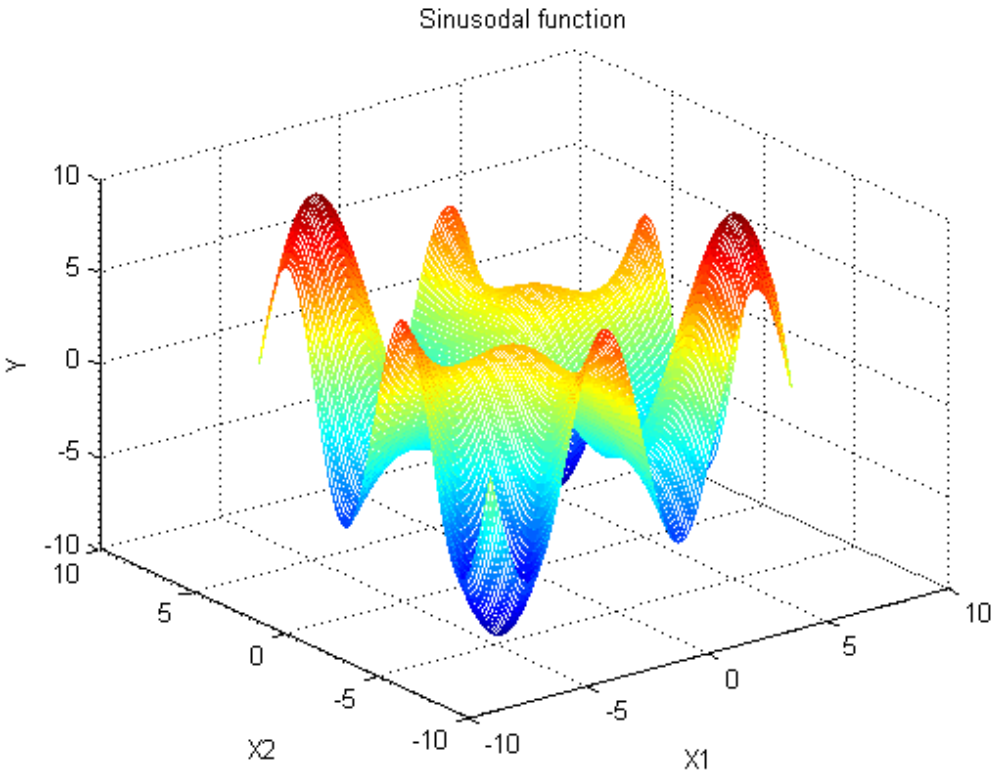


Figure 4.3: Two dimension sinusoidal test function.

caused by parameter estimation uncertainty is significant in the extrapolation case. Based on the results of the three numerical experiments, some conclusions can be made. The random noise $\varepsilon(x)$ inflates the overall prediction error as $\sigma_{\varepsilon}^{2}(x)$ increases. The additional prediction error caused by parameter estimation uncertainty becomes more important when the correlation is weak. This is aligned with the results and conclusions observed in Zimmerman & Cressie (1992b). Overall, the modified nugget effect model performs better than the nugget effect model and deterministic kriging model as observed in Yin *et al.* (2008). In the next step of this research, we can apply the kriging model with modified nugget effect to the actual experimental design and simulation optimization problem.

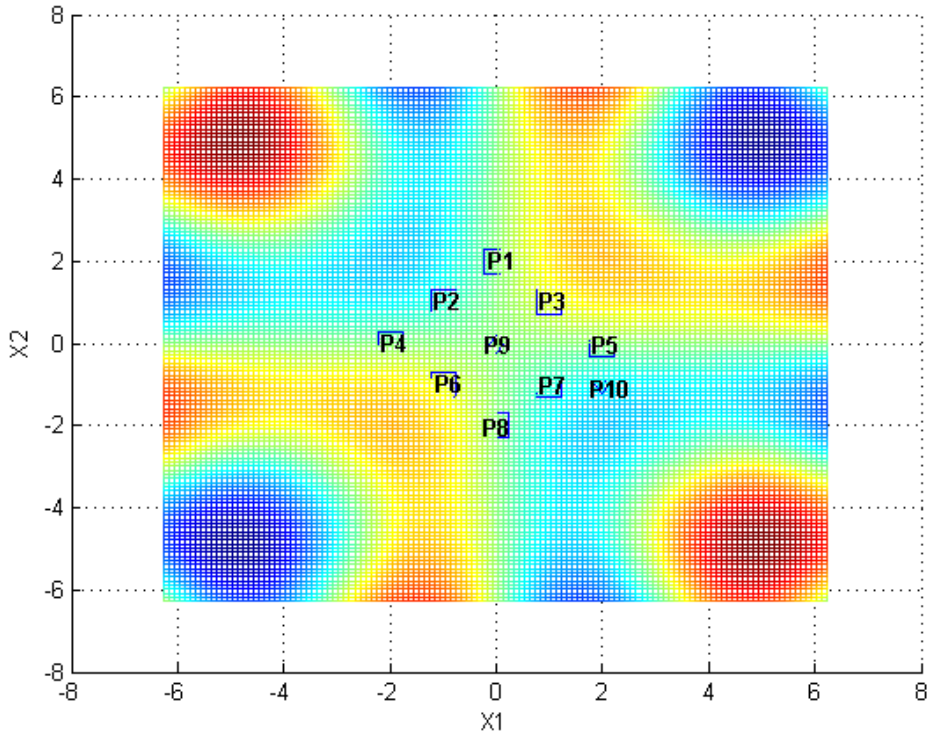


Figure 4.4: Design of the sinusoidal test function.

Table 4.6: Results for the two dimension linear test function.

Estimation based on noisy data						
		$\hat{\phi}_Z$				
		Mean	Var	ErrA	ErrN	ErrP
DK	P_9	118.2	5E3	0.199	0.191	0.005
	P_{10}	118.2	5E3	7.220	0.317	0.128
NK	P_9	49.81	4E3	0.192	0.180	0.004
	P_{10}	49.81	4E3	7.131	0.242	0.096
MK	P_9	16.26	2E3	0.148	0.129	0.002
	P_{10}	16.26	2E3	7.097	0.224	0.036
Estimation based on noiseless data						
		ϕ_Z^*			ErrM	
	P_9	14.39			0	
	P_{10}	14.39			6.72	

Chapter 5

OPTIMIZATION OF STOCHASTIC SIMULATIONS WITH KRIGING METAMODEL

5.1 Introduction

The development of the kriging model with modified nugget effect in Chapter 3 and Chapter 4 raises naturally the initial experiment design problem of properly allocating limited computing budget to optimize the objective function with a given system. In this chapter, we address this experimental design problem for the optimization of stochastic simulation models.

Some existing optimization methods for stochastic simulation, including the sequential Response Surface Method (RSM), Angün *et al.* (2002), the Stochastic Approximation (SA) method, see Kushner & Clark (1978), and some heuristic methods like the Genetic Algorithm and Simulated Annealing, are either insufficient for global optimization or computational expensive. In addition, Jones *et al.* (1998) proposed the Expected Improvement (EI) function and the Efficient Global Optimization (EGO) method to balance the local and global search for optimum of the unknown response surface as the solution to the global optimization of deterministic simulation model. The proposed method was a kriging

metamodel based optimization method developed from the Bayesian based optimization methods in Mockus *et al.* (1978) and Mockus (1994). The originally proposed EI function and EGO algorithm considered the allocation of the design points as the only design option for experimenter and focused on balancing the search for the local area of the current optimum and the entire sample space. However, for stochastic simulation, the random noise can considerably affect the experimental design, and the experimenter might also consider additional design option of adding replications to the design points to reduce the random variability other than only reducing the spatial uncertainty by observing new design points. Huang *et al.* (2006) extended the EGO scheme to Sequential Kriging Optimization (SKO) in order to be adaptable of the stochastic simulation model. With the nugget effect kriging model and augmented EI function, the SKO algorithm accounted for the influence of random noise $\varepsilon(x)$ and considered the new design option of adding replications on the existing design points other than exploration for new design points. However, the SKO only considered the homoscedastic cases, which suggest that the random noise function has constant variability throughout the entire sample space. For the more general case with heterogeneous variance, the SKO is not capable of capturing the behavior of the stochastic simulation model due to the misspecified assumption on the variance of random noise function.

In this chapter, we develop a new two-stage sequential framework that can be adopted for the optimization of stochastic simulation with heterogeneous variances. Similar to Jones (2001) and Kleijnen *et al.* (2011), we adopt the term optimization to mean a minimization or maximization, even when there are no constraints. The proposed new two-stage framework is able to correctly account for the influence of non constant variances in the design space, and hence balances the need to reduce spatial uncertainty and random variability with local and global search in a typical stochastic simulation scenario. This chapter differs from the work in Picheny *et al.* (2010) in that the noise variance function need not be known, and the proposed algorithm is less computationally demanding.

5.2 The expected improvement function

In the selection of search points for sequential based optimization algorithms, several criteria such as minimizing the response or minimizing a lower bound of the response, maximizing the probability of improvement or maximizing the expected improvement, have been proposed (see Jones (2001)). Of these search criterions, the improvement functions are most popular due to its ability to trade-off exploitation and exploration in order to converge to the global minimum. Here we review the improvement function initially proposed for deterministic simulations and an extension for stochastic simulations. Jones *et al.* (1998) defined the improvement function at any potential design point x_0 as $I(x_0) = \max[Z_{min} - Z_p(x_0), 0]$, where Z_{min} indicates the already observed (simulated) minimum of the mean function and $Z_p(x_0)$ is the random variable which accounts for the predictive uncertainty of the kriging predictor at design point x_0. At any unsampled design points, as the mean response $Z_p(x_0)$ is unknown, the improvement is averaged over the uncertainty in the response instead. The expected improvement is defined as

$$\mathrm{E}[I(x_0)] = \mathrm{E}[\max[Z_{min}(x) - Z_p(x_0), 0]] \tag{5.1}$$

This criterion is intuitive as it actively searches for design points that maximize the response improvement over the best, while considering also the uncertainty in the response at unobserved points. Based on the results from Jones *et al.* (1998), under the assumptions of the deterministic kriging model where $Z_p(x_0) \sim N\left(\hat{Z}(x_0), s_Z^2(x_0)\right)$, Equation (5.1) can be computed by

$$\begin{aligned} \mathrm{E}[\max[Z_{min}(x) - Z_p(x_0), 0]] &= (Z_{min}(x) - \hat{Z}\,(x_0))\Phi\left(\frac{Z_{min}(x) - \hat{Z}\,(x_0)}{s_Z(x_0)}\right) \\ &+ s(x_0)\phi\left(\frac{Z_{min}(x) - \hat{Z}\,(x_0)}{s_Z(x_0)}\right) \end{aligned} \tag{5.2}$$

Huang *et al.* (2006) adapted the EI function to address stochastic simulation systems. They adopted the nugget effect Kriging model (Cressie (1993)), where the random errors are assumed to be i.i.d. throughout design space, i.e. or all i,

to model the stochastic responses of the simulation and proposed the following augmented EI (AEI) function to drive the sequential search for the design points.

$$AEI[I(x_0)] = \mathrm{E}\left[\max[Z_{min} - Z_p(x_0), 0]\right] * \left(1 - \frac{\sigma_\varepsilon}{\sqrt{s^2(x_0) + \sigma_\varepsilon^2}}\right) \tag{5.3}$$

When x_0 is the current point of best response, the augmented factor is equal to the relative reduction in the prediction error at x_0 with the addition of another replication. The authors use this factor to represent the diminishing returns of additional replications at the current point of best response.

5.3 Limitations of EGO and SKO in Noisy Heteroscedastic Situations

Although applications of EGO (with EI) and SKO (with AEI) have been successful in deterministic and homogeneously random situations respectively, direct application to the heterogeneously random situations is not straightforward. To better understand the behavior of the algorithms and criterion, we first look at potential limitations of each in certain stochastic situations considered in this thesis. We first illustrate the limitations of EGO in a simple random situation. Consider the following example where the noisy test function is given as

$$Y(x) = Z(x) + \varepsilon(x) = (2x + 9.96)\cos(13x - 0.26) + \varepsilon(x) \tag{5.4}$$

where $\varepsilon(x)$ is a white noise function with zero mean, variance $\sigma_\varepsilon^2 = 2$ and $x \in [0, 1]$. The test function is sinusoidal with a slightly increasing gradient, with one local minimum at 0.2628, and a global minimum at 0.7460.

Starting with an initial 7 point Latin Hypercube Design (LHD) and 5 replications per point, the initial ordinary Kriging model fit and EI function are given in the left two plots of Figure 5.1. The main problem with fitting a deterministic Kriging model to stochastic simulation outputs is that the Kriging response interpolates the sample means, which can lead to a misleading estimate of the global minimum as seen inFigure 5.1. We see that even after 7 more input observations, the estimated response has mistakenly identified the wrong minimum point because of a low observation near the local minimum on the left. This problem

could be mitigated if there was a mechanism that could improve the precision of the sample means around the local minimums. Since the EGO algorithm lacks such a feature to deal with random variability, the algorithm is heavily penalized by a misleading initial fit. Applying the sequential SKO algorithm, we obtain

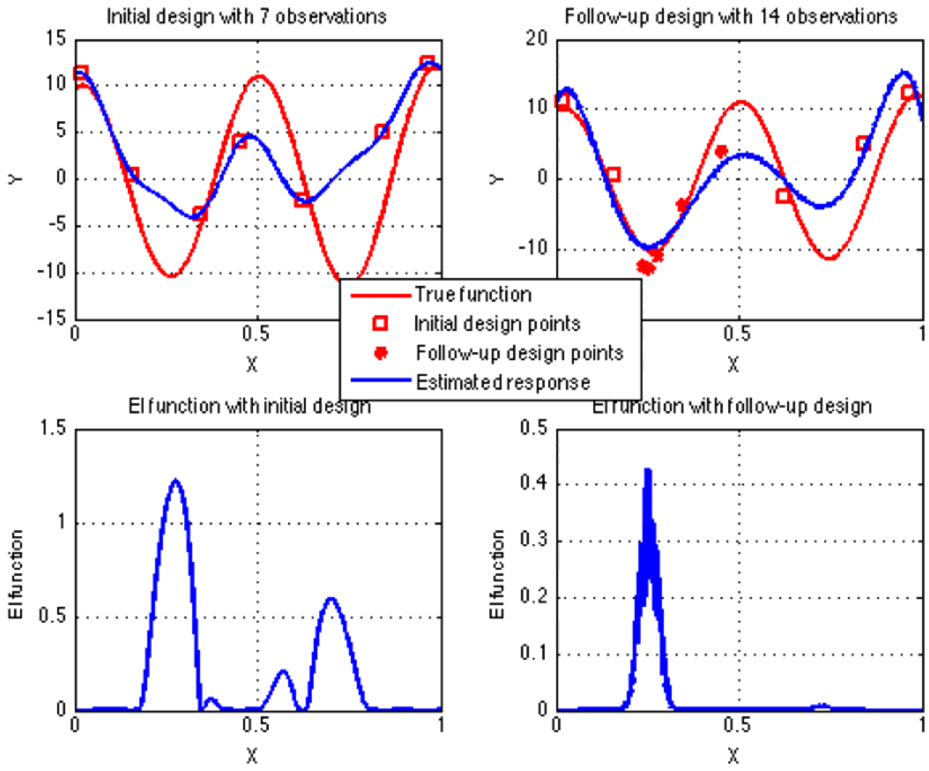


Figure 5.1: EI function and response metamodel with noisy test function (white noise)

a much better fit and results. Figure 5.2 displays the initial fit, AEI function and final response fit. As can be seen, with the appropriate metamodel form and sequential criterion, the results are much improved.

However, in the more general case of stochastic simulations with heterogeneous noise, the AEI function cannot properly account for the influence of random noise with non-constant variances. One straightforward method is to substitute the constant value in Equation (5.3) with the function $\sigma_\varepsilon(x)$. We modify the test function in Equation (5.4) so that the variance of the random noise now

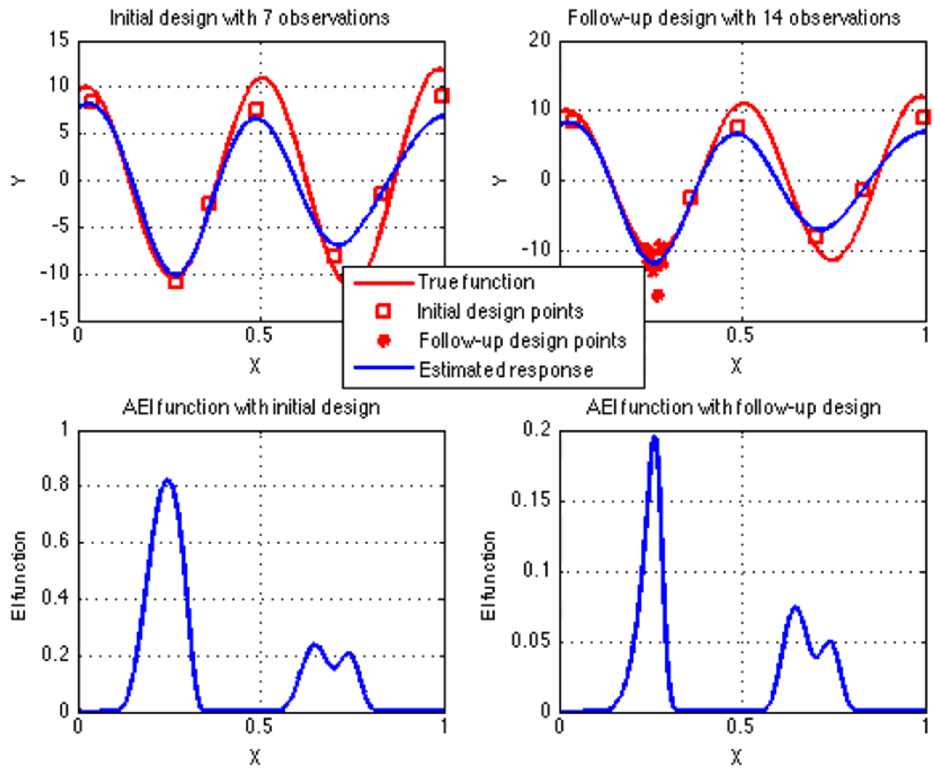


Figure 5.2: AEI function and response metamodel with noisy test function (white noise)

is a location dependent function with the linear form $\sigma_\varepsilon^2(x) = 1 + x, x \in [0, 1]$. Figure 5.3 displays the response model fit and the modified AEI function.

As seen in the upper left plot of Figure 5.3, the estimated response surface of the initial design has two local optima: the current best on the left hand side and a local sub-optimum on the right. Based on the functional form of variance function, the area on the left hand side has lower variability than the area on the right hand side. Hence in the follow-up sequential design, all the 7 new observations are located in the left region (area with lower noise and current best estimate). The AEI function still favors this left region after the follow-up design. This can be partially explained as the AEI function favors areas with lower variability due to the multiplicative effect of the augmented factor. Hence it is possible that the algorithm will be trapped in the local area with low variability for a long time. This is not a desirable property for a global optimization method,

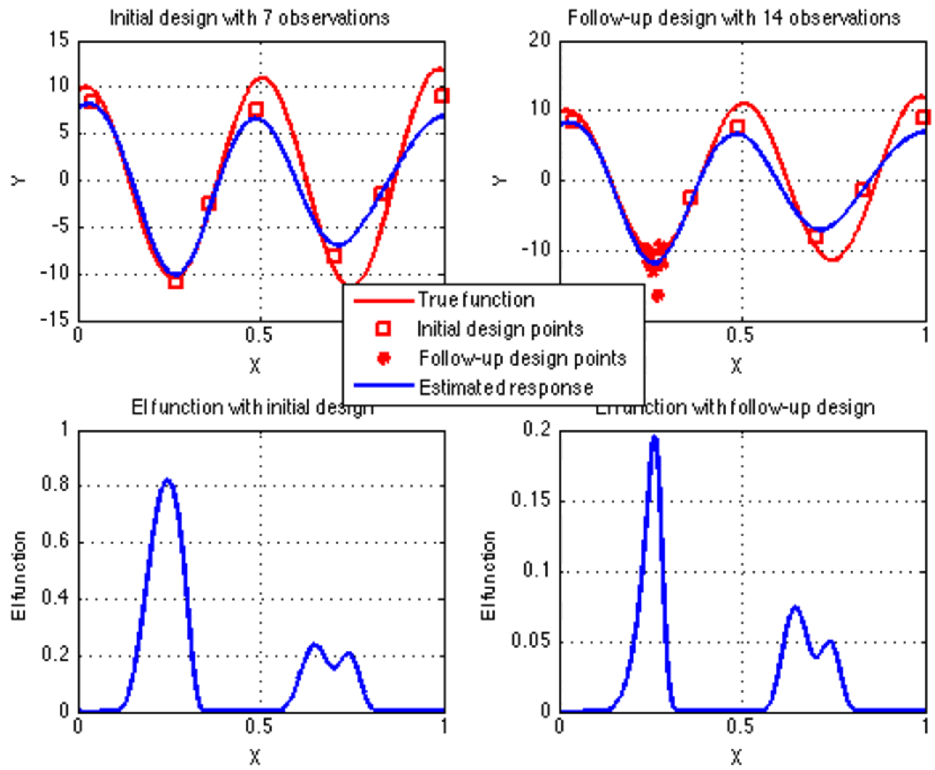


Figure 5.3: Modified AEI function and response metamodel with noisy test function (non constant variance)

especially when faced with limited budget constraints.

Another straightforward option is to simply use the original Expected Improvement function based on the MNEK model. The problem here is that the expected improvement value is less sensitive to decreases in predictor uncertainty $s(x)$ than $Z_{min} - Z^*$ as can be seen in Figure 5.4. This can lead to repeated selections of the current best design point at the expense of exploring other promising regions of the design space. The multiplicative factor in AEI was meant to address this issue by discouraging repeated selections of the current best point. However, applying the multiplicative factor to the case of heterogeneous noise would require that the noise variance function be known, and the sampling issue mentioned before will still persist.

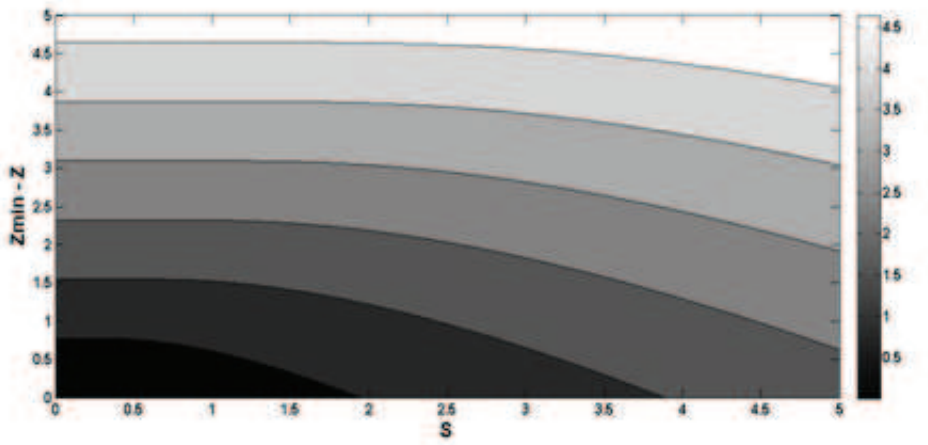


Figure 5.4: Contour plot of EI function of predictor mean difference and standard deviation using the Modified Nugget Effect Kriging model

5.3.1 Characteristics of Good Algorithms and Criteria

Based on the observations made above, we outline some challenges of adopting an EGO type framework for simulations with heterogeneous noise.

- The response model estimation is affected by both the location and computing effort at each design point. With heterogeneous variances and limited computing budget, the ideal distribution of computing effort is unlikely to be equal. An effective procedure should be able to make best use out of the limited computing budget in locating the global optimum.

- The EGO algorithm has the desirable characteristic of balancing exploitation and exploration. In order to retain this desirable characteristic in a stochastic environment, a new procedure should be able to search globally without having to exhaustively search a local region. Good estimates of Z_{min} are also necessary, especially in situations where there are several optima with values close to the global optimum.

- In situations with limited computing budget, it is useful for the algorithm to explore unexplored regions (improve chances of finding a better minimum) at the outset of the algorithm. However, as the budget is expended towards

the end, the focus should be fine tuning in the current best area as there will likely be insufficient remaining budget to find and differentiate a lower minimum elsewhere.

In the next section, we will detail the development of our search and allocation procedure that considers these desirable characteristics.

5.4 Development of Methodology

As previously noted, the accuracy and fit of the Kriging model can be drastically affected by random noise. It is therefore important to consider both the response and the noise levels carefully. As seen in Figure 5.4, direct application of the MNEK model may not have equitable effect on the EI criterion. The proposed methodology intends to address both considerations individually in a two stage sequential approach. This two stage iterative framework represents a division of labor approach towards the optimization of a stochastic objective function with heterogeneous noise variance. After the initial fit, each subsequent iteration of the algorithm is composed of a Search stage followed by an Allocation stage. In the first stage, denoted as the Search stage, the modified Expected Improvement (mEI) infill criterion is used to select a new point. In the second stage, denoted as the Allocation stage, the Optimal Computing Budget Allocation (OCBA) technique (Chen *et al.* (2000)) is applied to distribute an additional number of simulation replications among sampled design points. Specifically, the algorithms Search stage is responsible for identifying potential global optimum points, while the Allocation stage seeks to drive down uncertainty due to random variability at sampled points, with the goals of improving model fit at regions that contain local minimums and eventually correctly selecting the point of the global optimum.

The framework also features a division of allocation heuristic. Here we set the amount of computing budget per iteration as a constant, but the distribution of each iterations budget between Sampling and Allocation stages changes as the algorithm progresses. At the start, most of the iterations budget goes towards

exploration (search stage); towards the end, the emphasis is shifted to identifying the point of best response (allocation stage).

Together, this iterative two stage approach, along with the division of allocation heuristic, will sequentially search the design space for the global minimum while considering notably the fit of the surrogate model driving the search. The two stages of the framework will be described in detail in the next subsections.

5.4.1 The search stage

In this search stage, the following modified expected improvement function is proposed

$$mEI(x) = E\left(\max\left[Z_{min} - Z_p^*(x), 0\right]\right) \tag{5.5}$$

In this proposed criterion, Z_{min} is the predicted response at the sampled point with the lowest sample mean, and $Z_p^*(x)$ is a normal random variable with mean given by the MNEK predictor at x as shown in Equation (3.17) and variance given by the predictors spatial prediction uncertainty as shown in Equation (3.18).

The use of the MNEK predictor for the response is a straightforward choice that provides an unbiased prediction given the heterogeneous nature of the noise. What is different from a straightforward adaptation of the Expected Improvement criterion is the treatment of the predictor uncertainty. Since the allocation stage will by design address the stochastic noise, s_z^2 is used in this search stage only. This enables the search to focus on new points in promising regions with high predicted responses and new points that reduce the spatial uncertainty of the metamodel. In addition, by ignoring predictor uncertainty caused by random variability, s_ε^2, the modified Expected Improvement function assumes that the observations are made with infinite precision so the same point is never selected again. This allows the algorithm to quickly escape from a local optimum, and brings the sampling behavior closer to that of the original Expected Improvement criterion and its associated trade-off between exploration and exploitation.

5.4.2 The allocation stage

Since the Search stage is dedicated to the selection of a new point, the Allocation stage will have to manage random variability by intelligently allocating addi-

tional replications among sampled points. A related problem arises in Ankenman *et al.* (2010) work on Stochastic Kriging. In their work, the ultimate goal was to improve the global fit of the Stochastic Kriging response surface; therefore additional replications were distributed among sampled points with the objective of minimizing integrated mean squared error. In the case here where the ultimate goal is global optimization, additional replications are distributed with the goal of maximizing the probability of the correct selection of the sampled point x^{**} as the global optimum.

Alternative algorithms by Huang *et al.* (2006) and Picheny *et al.* (2010) identify the point with the best response (or global optimum) as the sampled point with the lowest predicted quantile response

$$x^{**} = \underset{x\in[x_1,x_2,\ldots,x_n]}{arg\,\min}\; Z(x) + cs(x)$$

where c is user defined in accordance with the user's risk tolerance, with a higher c corresponding to a lower level of risk tolerance and vice versa. For practical purposes, x^{**} in these algorithms are also optimized over previously observed locations.

Compared to the above method, OCBA formulates the global optimum selection problem as an optimization exercise and thus provides a more rigorous way of identifying the sampled point with the best response. OCBA uses the sample mean and sample variance as response and response uncertainty estimators respectively. As the allocation stage aims to improve the data used to fit the model, especially in promising regions, adopting these estimators at this stage is reasonable.

Assuming that we have n sampled points, with each point x_i having a sample mean given by $\bar{Y}_i$, sample variance $\sigma_\varepsilon^2(x_i)$ and replication number R_i, then according to Theorem 1 provided by Chen *et al.* (2000), the Approximate Probability of Correct Selection (APCS) can be asymptotically maximized when available computing budget tends to infinity and when

$$\frac{N_i}{N_j} = \left(\frac{\sigma_\varepsilon(x_i)/\delta_{b,i}}{\sigma_\varepsilon(x_j)/\delta_{b,j}}\right)^2, \quad i,j \in 1,2,\ldots,n, \text{and} i \neq j \neq b \tag{5.6}$$

$$N_b = \sigma_b \sqrt{\sum_{i=1, i \neq b}^{k} \frac{N_i^2}{\sigma_i^2}} \tag{5.7}$$

where N_i is the number of replications allocated to point x_i, x_b is the point with the lowest sample mean and $\delta_{b,i}$ is the difference between the lowest sample mean and the sample mean at point x_i. Adopting this allocation rule to maximize the APCS, at the end of the Allocation stage, the sampled point with the lowest sample mean will be selected as the location of the best response. The estimate of the best response, Z_{min}, which will be used in the mEI criterion will be given by the MNEK predictor.

An additional benefit of using the OCBA technique is that the majority of the additional replications will be allocated to points with low sample means and high sample variances in order to separate the best point from the contenders. Assuming that the lowest sample means lie close to local minimums, then OCBA can utilize the additional replications to tighten the metamodel fit at regions around the local minimums. This in turn can help the modified Expected Improvement criterion make a better selection in the subsequent iteration.

5.4.3 An algorithm overview

In the proposed framework, the procedure alternates between the Search and Allocation stages until the budget is consumed. As discussed, in such an approach, it is desirable to start off with more exploration and end off with exploiting the best few minimums identified. Here, we adopt a simple division of allocation heuristic to divide the budget allocation to the different stages on the proposed framework.

Here $r_A(i)$ can have a general form as $r_A(i) = g(T, B, r_{min}, i)$, and hence $r_S(i) = B - r_A(i)$. According to the characteristics of good algorithm in Section 5.3.1, $r_A(i)$ should be a monotonic increasing function of i, which allows the algorithm to focus on searching for new point in the beginning and tuning current best area in the end. $r_A(i)$ should also be a increasing function of $B - r_{min}$, which suggest that we could add extra efforts to the Allocation stage if the number of

Table 5.1: Algorithm parameter definitions.

Parameter	Definition
T	Total number of replications at the start
B	Number of replications available for each iteration
n_0	Size of initial space filling design. $n_0 B \leq T$
i	Current iteration. $i = 0, 1, 2, \ldots, I$. $I = \lceil \frac{T - n_0 B}{B} \rceil$
l	The accuracy limit for MEI value in Search Stage
r_{min}	Minimum number of replications for sampling a new point. $r_{min} \leq B$
$r_A(i)$	Number of replications available for Allocation Stage at iteration i
$r_S(i)$	Number of replications available for Search Stage at iteration i

available computing budgets for every single iteration B is much larger than the minimum computing budget for single observation point r.

Before the proposed algorithm can begin, the user needs to set the parameters T, B, n_0, l and r_{min}. T will be constrained by practical considerations such as the total time available for the entire optimization exercise and the average length of a simulation run. The size of the initial design, n_0, can be set to $10d$ as suggested by Jones *et al.* (1998) where d is the number of dimensions. This $10d$ recommendation has been extensively studied by Leoppky *et al.* (2009) and found to be a reasonable rule of thumb for initial designs. l can control the accuracy of the Search Stage, we can use the recommendation of $l = 0.015$ from Locatelli (1997). length o r_{min} and B should be set such that there are sufficient replications available for the first Allocation stage. As a rough guide, B should be set such that the number of replications available for the first Allocation stage is at least 5 times the number of design space dimensions. An algorithm can be proposed as follows.

- Step 1: Initialization: Run a size n_0 space filling design, with B replications allocated to each point.

- Step 2: Validation: Fit a MNEK response model to the set of sample means, and use cross validation to ensure that the Kriging prediction is satisfactory.
- Step 3: Search for the location x_0 that maximize the modified Expected Improvement criterion Equation (5.5).
- Step 4: Compute the computing budgets for the Allocation Stage and Search Stage, proceed to the each Stage accordingly. Initailize the parameters, set $i = 1, r_A(0) = 0$. If $mEI(x_0) < l$, go to Step 5.
- Step 4a: While $i \leq I$
  $r_A(i) = \min(\lfloor (B - r_{min}) \left(\frac{i}{I}\right)^{\frac{I}{n_0}} \rfloor, T - n_0 B)$
  If$(T - n_0 B - (i-1)B - r_A(i)) > 0$, $r_S(i) = B - r_A(i)$ end
- Step 4b: Sample a new point at location x_0 with $r_S(i)$ replications. (Search Stage)
- Step 4c: Using OCBA (Equation (5.6) and Equation (5.7)), allocate $r_A(i)$ replications among all sampled design points. (Allocation Stage)
- Step 4d: Fit a MNEK response surface to the set of sample means. $i = i+1$, if $i > I$, the algorithm ends, else compute the mEI function and update the location x_0 that maximize the mEI function. If $mEI(x_0) < l$, go to Step 5, else go to Step 4a.
- Step 5: Using OCBA to allocate all the remaining computing budget $T - (i-1) * B$ to the existing points, the algorithm ends.
- The point of the global optimum at the end will be the point with the lowest sample mean.

As the starting parameters which determine the number of iterations and the computing budget used per iteration are set prior to collecting any data, there is a possibility that the recommended parameter settings are unsuitable for the problem. In step 2, the cross validation test can provide feedback regarding the suitability of the initial parameters T, B, n_0, l and r_{min}. Cross validation is a statistical technique that can assess the generalization of statistical analysis for

independent data set. A general introduction to cross validation can be found in Mosteller & Tukey (1968). Cross validation can be used to check the accuracy of the predictive model. For this kriging metamodel application in stochastic simulation, ff a single design point or more fails the cross-validation test, it could mean that the current amount of computing budget is insufficient for dealing with the noise in the response. Possible remedies include increasing B, or increasing the number of design points around the point(s) that fail the cross-validation test, or to apply a log or inverse transformation to the response as suggested by Jones *et al.* (1998).

After successful validation, the algorithm will approach to the Search and Allocation stage with the remaining computing budgets. According to the characteristics of good algorithms in Section 5.3.1, the computing budget set aside for the Allocation stage $r_A(i)$ should increase as the algorithm approaching to the end, which suggests that $r_A(i)$ should be a monotonically increasing function of current iteration number i.

Moreover, the selection of initial design size n_0 will determine the size of following-up design given a fixed T and B, which will be expressed in terms of $I = \lceil \frac{T-n_0B}{B} \rceil$. For different I, the $r_A(i)$ should have different increasing rate in order to meet the requirement of distributing more replications to the Allocation stage when $i \to I$. For example, considering the case with $T = 320$, $B = 40$, the total design size is $T/B = 8$. The first scenario with initial design size $n_0 = 4$, the total iteration number for following design is $I = \lceil \frac{T-n_0B}{B} \rceil = 4$. The second scenario with the initial design size n_0 to 6, the total iteration number for following design will be $I = \lceil \frac{T-n_0B}{B} \rceil = 2$. Considering $r_A(2)$, the number of replications for Allocation stage for iteration 2, it will be the middle iteration of the following-up design for the first scenario and the last iteration for the second scenario. Hence the $r_A(2)$ in the second scenario should be larger than it in the first scenario to maintain the characteristic of good algorithm.

As a result, we propose to use formula for $r_A(i)$ in Step 3a. First, $r_A(i) = \min(\lfloor (B - r_{min}) \left(\frac{i}{I}\right)^{\frac{I}{n_0}} \rfloor), T - n_0B)$ is a monotonically increasing function of i, which gives the algorithm a desirable characteristic of placing more emphasis on exploration at the start, and focusing more on exploitation at the end. Second, $r_A(i)$ is a function of ratio $\frac{I}{n_0}$, which can automatically adjust the increasing rate

of $r_A(i)$ given the ratio $\frac{I}{n_0}$, which is a function of initial parameters $T, B, \text{and} n_0$. Considering a simple example with $T = 400, B = 40$, and $n_0 = [3, 4, 5, 6, 7]$ for 5 different scenarios. the ratio $\frac{I}{n_0}$ will be $\frac{I}{n_0} = [2.3, 1.5, 1, 0.67, 0.43]$, accordingly. The changes of $r_A(i)$ in terms of iterations for different scenarios are given in Figure 5.5.

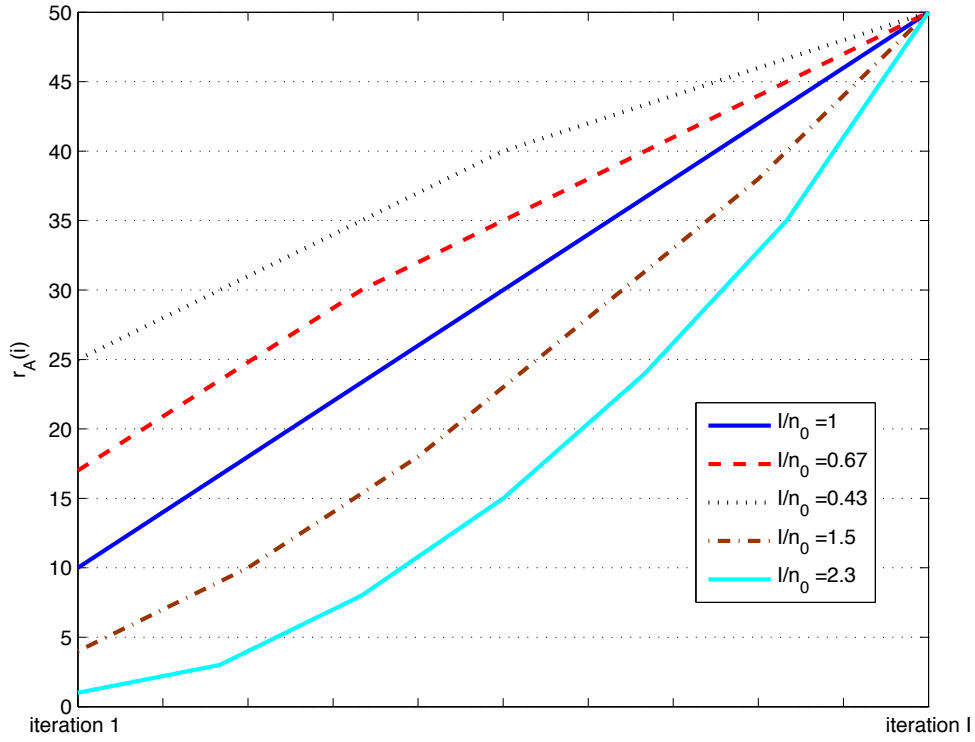


Figure 5.5: $r_A(i)$ at different iterations as $\frac{I}{n_0}$ changes.

As can be seen, the initial parameter set with lower $\frac{I}{n_0}$ ratio will distribute more computing budgets as the additional replications to the Allocation stage. This behavior is reasonable as a low $\frac{I}{n_0}$ ratio indicates the following-up design is close to the end of algorithm, which suggests that the algorithm should focus on the improvement of current best locations. In a special case with large size of initial design , $\frac{I}{n_0} \to 0$, $r_A(i) \to B - r_{min}$, which means the algorithm try to distribute maximum number of replication for the Allocation stage regardless of

iteration number.

Moreover, the parameter l is used to control the accuracy of the mEI function. This parameter becomes important when the total computing budget T is huge. If $T \to \infty$, then $I \to \infty$, which suggest that the total number of design points goes to infinity. According to the results in Jones *et al.* (1998) and Locatelli (1997), we can show that the proposed algorithm can converge to the optimal in probability. See Appendix D for details.

In summary, the proposed algorithm relies on two techniques to handle the tasks of selecting new points and managing the heterogeneous nature of the noise. From a modeling viewpoint, we can relate the algorithm back to the Kriging model form and its overall iterative goal of optimization. The selection of a new point in the Search stage can be seen as improving the fit of MNEK predictor $\hat{Z}(x_0)$ for the subsequent iterations by improving the estimation of β_Z (see Equation (3.17)) and reducing the uncertainty in the prediction ,especially for promising regions of x_0. The purpose of the Allocation stage is twofold. This stage first improves the fit of $\hat{Z}(x_0)$ by improving the estimation of β_ε (see Equation (3.17) and Equation (3.18)) and reducing the uncertainty in the prediction (last term in Equation (3.17)). Secondly, the resource allocation goal of maximizing the probability of correct selection also improves the estimation of Z_{min} for the Search stage in the next iteration.

5.5 Numerical Examples

To better illustrate the behavior of the proposed algorithm, numerical example on the single dimension function is conducted. In the second example, the performance of the proposed varied rate algorithm is compared with Picheny *et al.* (2010)'s approach, which is denoted as EQI in the subsequent sections.

5.5.1 Single dimension test function(Comparative study)

We applied the proposed algorithm to the one dimensional example in Section 3.3 with the noise variance function $3(1+x)^2$. A Latin Hypercube design was

used for the initial fit for both algorithms and the following starting parameters were adopted:

Table 5.2: Example starting parameters.

Parameter	Value
T	320
B	40
n_0	5
l	0.015
r_{min}	10

Under current test function setting, the newly added points and computing budgets each stage for the proposed algorithm are listed in Table 5.3 (Here SSB stands for computing budgets distributed for Search Stage and ASB stands for computing budgets distributed for Allocation Stage).

Table 5.3: computing budgets and location each iteraion for the proposed algorithm.

Iteration	SSB	ASB	New design location
1	30	10	0.7380
2	10	30	0.7330
3	10	30	0.2660

We can change the starting parameters to simulate a scenario with a small initial design size as listed in Table 5.4.

As seen in Figure 5.7 and Table 5.5, for the case with smaller initial design size, the proposed algorithm focus on the Search stage in the beginning by assigning most of the replications to observe a new point. This behavior allows the algorithm to observe new points with low variability in the early iterations.

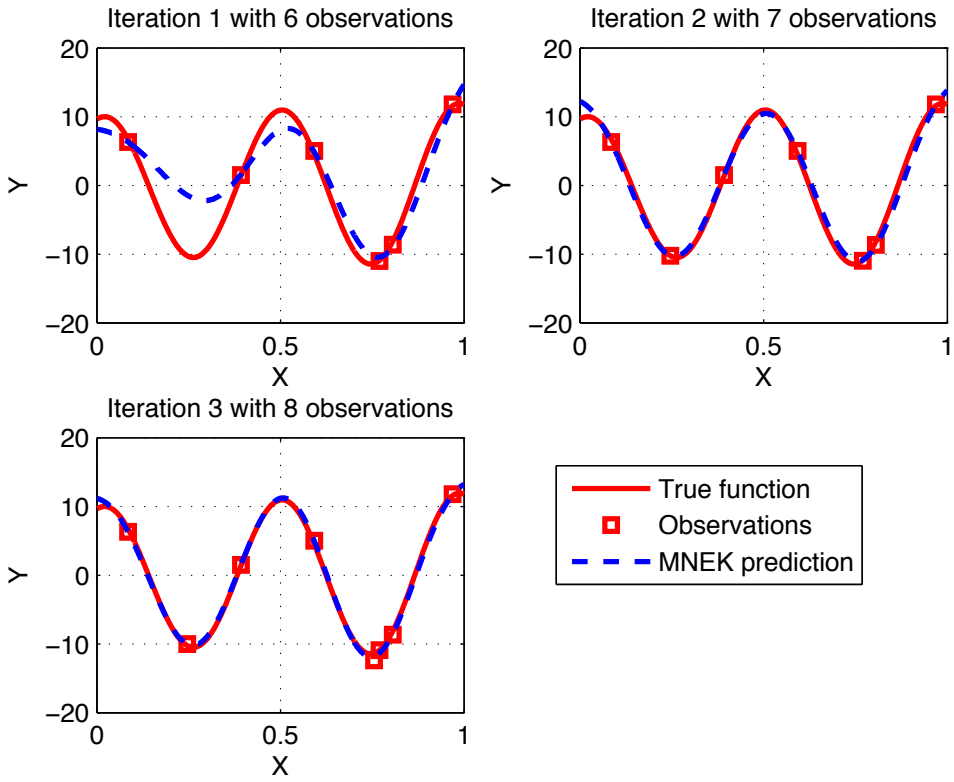


Figure 5.6: One dimensional example with proposed algorithm.

Hence the algorithm is able to obtain a satisfactory response with a few iterations and focus on improving the promising region in the end.

5.5.2 Two Dimension Keys and Reese (2004) Function (Comparative Study)

Like the proposed algorithm, EQI was developed for optimization of simulations with heterogeneous noise. Out of the two resource allocation schemes in EQI, we selected EQI with online allocation for comparison because it is less computationally demanding than the alternative. Furthermore, EQI with online allocation also has a two-stage approach that first selects a new point that maximizes the EQI criterion, followed by the allocation of additional computing resources to the point until a criterion is satisfied.In the rest of the section.

Table 5.4: Example starting parameters(small initial design size).

Parameter	Value
T	320
B	40
n_0	3
l	0.015
r_{min}	10

Table 5.5: computing budgets and location each iteraion for the proposed algorithm. (small initial design size)

Iteration	SSB	ASB	New point location
1	38	2	0.3120
2	34	6	0.6940
3	28	12	0.6450
4	20	20	0.7500
5	10	30	0.7590

The test function used for the comparison is a modified version of the tetramodal function used by Keys & Rees (2004):

$$\begin{aligned} Z(x_1, x_2) &= -5(1-(2x_1-1)^2)((1-(2x_2-1)^2))(4+2x_1-1)(0.05^{(2x_1-1)^2} \\ &- 0.05(2x_2-1)^2)^2 \end{aligned}$$

Both dimensions of the test function, x_1 and x_2, are scaled to $[0, 1]$. The test function's contour plot is shown in Figure 5.8. The global minimum is located at $[0.85, 0.5]$ and has the value -7.098. The test function was overlaid with a noise function with standard deviation given by $1.2x_1$. The comparisons between the two algorithms were carried out for 100 macro-replications. The variance of the noise in EQI was taken from the actual function while the proposed algorithm

was run under two settings - one with the noise variance known, and the other with the noise variance estimated using sample variances. In every comparison, both algorithms started off with the same size 20 LHS design with 40 replications at each point and the same set of initial sample means. The 50th percentile was used in the EQI criterion. Other algorithm parameters are shown in Table 5.7. In EQI, T_{20} refers to the post initial fit budget and γ is a user defined parameter

Table 5.6: Algorithm's setting.

Proposed Algorithm		**EQI**	
Parameter	**Setting**	**Parameter**	**Setting**
$T - n_0 B$	200	T_{20}	200
B	40	t_e	10
r_{min}	10	γ	0.5

with value from 0 to 1 that controls how much available budget is allocated to the newly sampled point. t_e refers to the "step size", which is the number of replications allocated to an existing point or used to sample a new point with every EQI action. γ was set to a value quoted by Picheny *et al.* (2010) while a step size of 10 was chosen to match r_{min} in the proposed algorithm.

The algorithms were first evaluated on how close they were able to get to the true location and value of the global minimum. Table 5 shows the results. Statistically, there was no significant difference between the 3 sets of results. The proposed algorithm was able to match EQI's performance in both evaluation criteria. For this example, it showed that the proposed algorithm was able to achieve a level of accuracy similar to EQI. Furthermore, the proposed algorithm achieved it with less computational effort than EQI. Taking the number of matrix inversions as a measure of computation effort required, we saw that the proposed algorithm required 5 inversions in total per macro-replication - 1 for the initial fit, and 4 more for the subsequent iterations. As for EQI, there were a total of 200001 inversions per macro-replication. With the design space discretized into a 100 by 100 grid,the EQI value at every one of these grid points would have to

Table 5.7: Example computing budget distribution.

	EQI		**Proposed Algorithm**		**Proposed Algorithm**	
	(Noise Variance Known)		(Noise Variance Known)		(Noise Variance Estimated)	
Evaluation Criteria	Average	Std.[1]	Average	Std.	Average	Std.
$\lvert x^{**}_{pred} - x^{**}_{true}\rvert$	0.318	0.253	0.318	0.249	0.312	0.247
$\lvert \hat{Z}(x^{**}_{pred}) - Z(x^{**}_{true})\rvert$	1.669	0.998	1.652	0.991	1.656	0.990

be evaluated which would involve the inversion of a $n+1$ by $n+1$ matrix, where n is the number of sampled points.

Another advantage of the proposed algorithm over EQI is that the noise variance function need not be known. In cases where the noise variance function is unknown, EQI requires that the noise variance function be estimated. Unlike EQI, the proposed algorithm has no such requirement as it uses sample variances as estimates of the variance of the noise at sampled points.

On the other hand, the proposed algorithm's division of computing resource heuristic is not as flexible as EQI's online allocation scheme. Although the proposed algorithm is able to adaptively allocate replications among sampled points, the proposed algorithm's computing budget for every iteration and the total number of iterations are fixed prior to the start of the algorithm. In EQI's case, the computing resource distribution between its sampling and replication allocation stages is "online "rather than pre-determined, so the resource distribution scheme makes use of increasing information as the EQI algorithm progresses while accounting for the diminishing computing budget. Other allocation schemes where the budget for each iteration $B(i)$ is dynamically determined can be explored for our algorithm and this is an area for future research.

[1]Std. stands for standard deviation

5.6 Ocean Liner Example

In recent years, the increasing bunker prices have threatened the shipping liner companies' accounting bottom line. To compensate for increasing prices, over two hundred shipping companies have adopted the strategy of slow steaming, especially in long-haul loops like Asia to Europe and Asia to North America. From empirical estimations, lowering the cargo vessel speed by 20% can save up to 50% of daily fuel consumption (Notteboom & Vernimmen (2009)). Although slow steaming results in having to increase the number of vessels within the service to maintain a satisfactory service level, this additional capital cost is usually offset by the savings from fuel costs. Another challenge facing shipping companies is deciding where to refuel (bunker) along a given route and the amount to bunker at these selected ports.

A recent research study by Yao *et al.* (2011) modeled the shipping fuel management planning problem as a mixed integer linear program. The linear program minimizes bunker fuel related costs by determining optimal ship speeds between ports, and selecting bunkering ports and bunkering amounts. Based on real data obtained from a shipping company, they empirically modeled the relationship between fuel consumption rate and ship speed for different ship sizes and expressed the daily fuel consumption rate as $F(V) = k_1 V^3 + k_2$, where k_1 and k_2 are two constants that depend on ship size and V is ship speed in knots.

This previous work can be extended to consider the operational level problem by considering more realistic operational conditions. This includes accounting for fuel consumption rate uncertainty and bunker price variation. With random fuel consumption, the ship's bunker fuel inventory safety level becomes an important system level decision as a fuel stock out scenario is possible and can be very costly to liner operations. In addition, as the distances between each port within the service can differ by thousands of nautical miles, safety levels at each leg should differ, each reflecting about equal risk of a fuel stock out within the leg. Due to the complexities of the real operational system and impact of the random components, to determine the safety levels for a given shipping service, a simulation optimization approach is applied to a rolling horizon discrete event simulation model.

In the rolling horizon discrete event simulation, the following takes place upon the arrival of a ship at each port,

- Observe the time taken and fuel consumed from the previous leg.
- Update the bunker inventory level based on the observed consumption. Update bunker prices at all ports based on current market prices.
- Solve mathematical optimization problem to obtain the optimal fuel management decisions for the bunkering ports, bunkering amounts and ship speeds.
- Bunker fuel amounts at current port (if necessary) and set ship speed for next leg.

This procedure yields the bunkering amounts, ports and ship speeds one port at a time, with each arrival in the simulation. Incorporating the mixed integer linear program in the rolling horizon discrete event simulation model facilitates a more realistic operational representation of the bunker fuel management problem, and enables better evaluations of system strategies (such as bunker safety levels).

Here we consider the Asia-Europe Express (AEX) service route of a shipping company, see Figure 5.9. As each run of the simulation incorporates the rolling horizon mathematical optimization model to determine the optimal fuel management strategy, each run is quite time consuming, taking around 5 minutes on an Intel Core 2 Duo 2.4 GHz computer. Consequently, one must carefully allocate the simulation budget to efficiently and effectively search the multidimensional decision space for the optimal safety levels.

We applied our proposed algorithm to determine the bunker inventory safety levels for the AEX service. Similar to Yao *et al.* (2011), the overall objective for this problem is to minimize the total bunker fuel related costs. We adopt the cost model in Yao *et al.* (2011) with an additional penalty function to describe the impact on the liner operation if a fuel out situation occurs at sea. Here we set the penalty function to be the product of the bunker capacity, bunker price at the next port and a penalty factor. We set the penalty factor value at 2. This reflects a high approximate cost of sending out an additional vessel to fill up

at sea, and can be adjusted based on individual company's costs and penalties. For example,if a ship with a bunker capacity of 3000 tons ran out of fuel before the port Yantian (where bunker costs $460 per metric ton), the penalty cost will amount to $3000 \times 460 \times 2 = 2,760,000$.

To account for the actual consumption rate variation, a noise term is added to the original empirical model: $F' = k_1 V^3 + k_2 + \varepsilon(V)$. Here we also assume that noise ε is a function of the ship speed. In addition, based on real data obtained from a shipping company, we observe the following three characteristics:

- The noise term follows a normal distribution with a zero mean and a standard deviation that increases with ship speed.
- The coefficient of variation CV, the ratio of standard deviation to mean daily bunker consumption rate, is a constant.
- This constant is different for different sized ships, where the bigger the ship, the smaller the number.

Hence the noise function is a normal distribution, $N(0, (CV * F(V))^2)$. Bunker prices at each port were obtained from Bloomberg for the period of Jan - July 2009. The prices were estimated to follow a normal distribution with means given in Table 5.8. The variances were also estimated from the same set of data, and as they were approximately similar for all ports, the same variance, $\sigma^2 = 15$, was used for all ports.

For the first test scenario, we compared our proposed approach with the results of a brute force search. A small computing budget of 80 was used, and we assumed a single bunker inventory safety level for all 15 legs of the route to facilitate the feasibility of the search. Table 7 provides relevant parameters for the AEX service obtained from the shipping company, some of which were estimated from the company's data.

Here the bunker inventory safety level was defined as the percentage of the total bunker capacity (3000 metric tons).For the brute force search, we observed 16 evenly distributed points from 0.5% to 10%, with 5 replications at each point. Table 8 provides the optimal safety levels obtained by the two methods, and the expected optimal costs were estimated using an additional 50 replications

Table 5.8: Bunker price parameters for ports in AEX route.

Ports	Mean Price $ ton
Hakata	460
Kwangyang	463
Pusan	468
Shanghai	457
Kaoshiung	463
Hong Kong	467
Yantian	458
Singapore	460
Rotterdam	461
Hamburg	461
Thamesport	456
Colombo	462
Singapore	459
Hong Kong	461
Kaoshiung	459

conducted at the optimal levels. From the table, we see that with the same computing budget, our proposed approach is able to obtain a lower cost solution. To conduct a more comprehensive brute force search, we observed 51 evenly distributed points in the same region, each with 5 replications. The last row of Table 8 shows the optimal results with the expected optimal costs computed based on 50 replications. Comparing our approach with the more comprehensive brute force search, we see that our proposed approach was able to obtain a solution close to the solution from the comprehensive search with less than 35% of the expense.

Overall, the results suggest that the proposed two-stage approach can more effectively distribute the computing budget to promising regions, finding better solutions with a fixed budget, or finding equivalent results with much less budget.

Table 5.9: Parameters for ports in AEX route.

Parameters	Values
Bunker capacity	3000 tons
Number of port calls	15
Coefficient of variation	0.07
Variance of bunker price	15 $\$^2 ton^2$
k_1	$0.007297 tons(day * knots^3)$
k_2	71.4 tonsday
Ship size	5000 TEU

This can be very useful for the problems with limited computing budgets.

To more realistically solve this problem, we relaxed the assumption on the bunker inventory safety level and considered different safety levels for different legs. This is more reasonable as the distances for each leg within the service can differ by thousands of nautical miles. With 15 different legs within the service, we considered this as a 15 dimension decision problem. A computing budget of 3000 replications was used, which is a reasonable amount reflecting the typical limited time decision makers have in the light of changing prices and short dynamic time horizon. 2000 replications was spread evenly among a $n_0 = 200$ initial space filling LHD, and the remaining were allocated according to our proposed approach described in Section 4.3. Table 9 shows the results.

As expected, bunker safety inventory levels increased with increasing distances between port calls. With varying safety levels, the optimal total cost was lower than the single safety level at 1.01341E+8.An interesting observation is that the increase in safety levels was not linear - the increments in safety levels decreased as distances increased. This was mainly due to bunker fuel management decisions in each horizon. As observed in Yao *et al.* (2011), the time windows of the shipping schedule have a major impact on the speeds of each leg and hence the bunkering decisions. As observed in the simulation, the resulting bunkering ports were usually ports before a long leg.As the optimal bunkering amounts at

Table 5.10: Optimal results for single bunker inventory level.

	Expected optimal cost	**Optimal minimum bunker inventory safety level**	**Total replications required**
Two-stage approach	$ 1.14304E+8	5.55%	80
Bruth-force search (16 points design)	$ 1.14309E+8	5.49%	80
Bruth-force search (51 points design)	$ 1.14301E+8	5.60%	255

these ports were generally high (due to the long leg ahead and quantity based discounts considered), the influence of the safety levels was reduced. Moreover, ships speed up (or slow down) based on the time windows of the schedule, thereby increasing (decreasing) consumption. It was also observed that legs with tight time windows had on average higher travel speeds, thus requiring higher safety levels. The results here clearly illustrates the interrelatedness of operational level bunker fuel management decisions and system level safety level decisions, and the necessity to consider both together in a stochastic simulation environment to ensure the robustness of the decisions.

5.7 Conclusion

In this chapter, we proposed a two stage sequential framework for the optimization of stochastic simulations with heterogeneous variances. The proposed two-stage framework is based on the Kriging model and iteratively incorporates the optimal computing budget allocation techniques and the modified expected improvement function to drive and improve the estimation of the global optimum. We first illustrate our approach with several numerical examples. The empirical results indicate that the proposed approach is effective in obtaining the optimal solutions and require less computer time than other Kriging based optimization techniques

Table 5.11: Optimal results for 15 different legs.

Port legs	Distance(NM)	Optimal bunker inventory safety level
Hakata - Kwangyang	171	1.2%
Kwangyang - Pusan	86	1.1%
Pusan - Shanghai	475	2.3%
Shanghai - Kaoshiung	639	2.4%
Kaoshiung - Hong Kong	350	2.3%
Hong Kong - Yantian	45	0.9%
Yantian - Singapore	1500	3.2%
Singapore -Rotterdam	8339	7.4%
Rotterdam - Hamburg	223	1.6%
Hamburg - Thamesport	325	1.9%
Thamesport - Colombo	6727	7.1%
Colombo - Singapore	1567	3.2%
Singapore - Hong Kong	1460	3.1%
Hong Kong - Kaoshiung	350	2.0%
Kaoshiung - Hakata	904	2.7%

proposed to address stochastic simulations. We also applied the approach to a real complex ocean liner simulation model to determine the optimal bunker inventory safety levels for a fuel management problem. The results from this problem provided invaluable insights on the inventory levels on a service route, and clearly illustrated the interrelatedness of the bunker fuel management and safety levels.

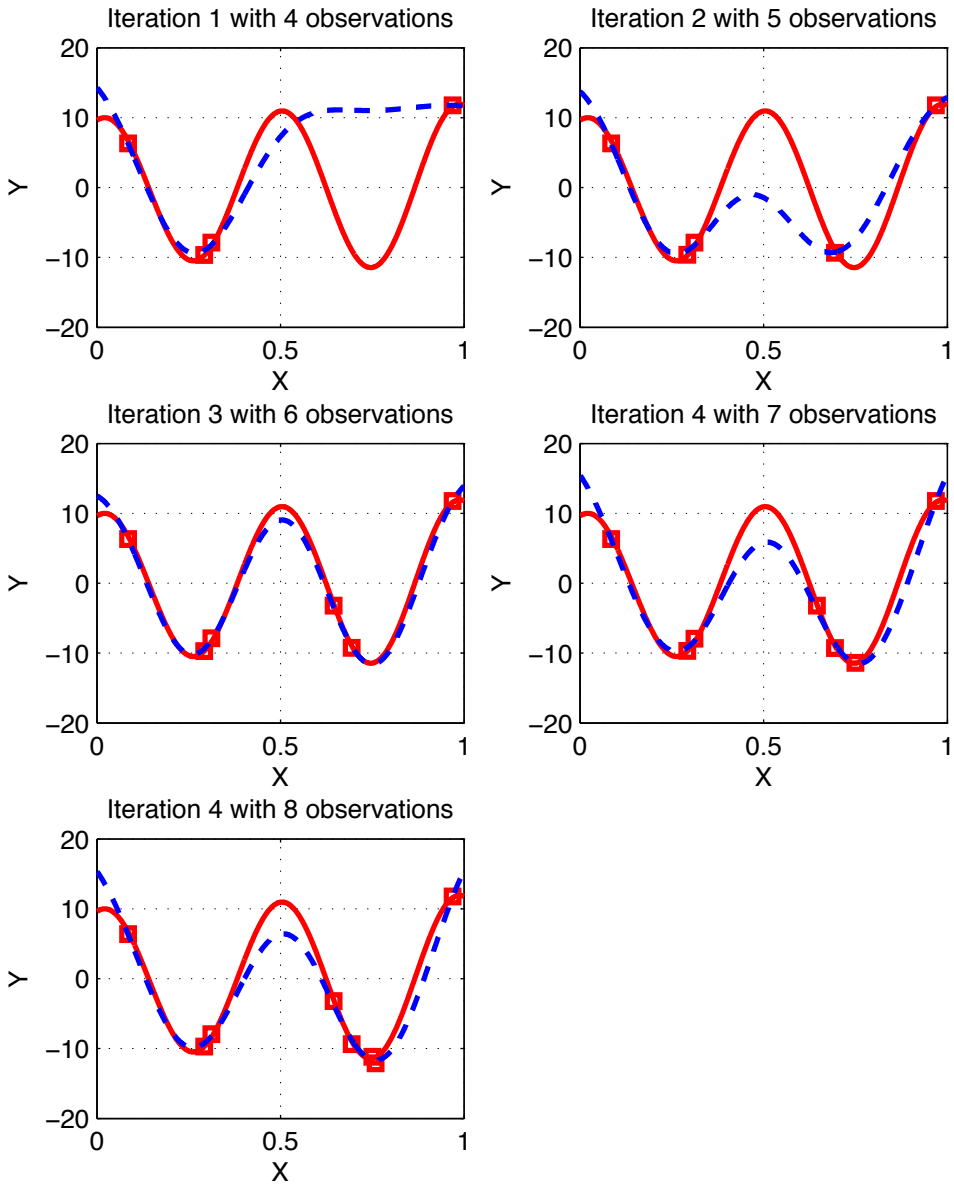


Figure 5.7: One dimensional example with proposed algorithm.(Higher variance)

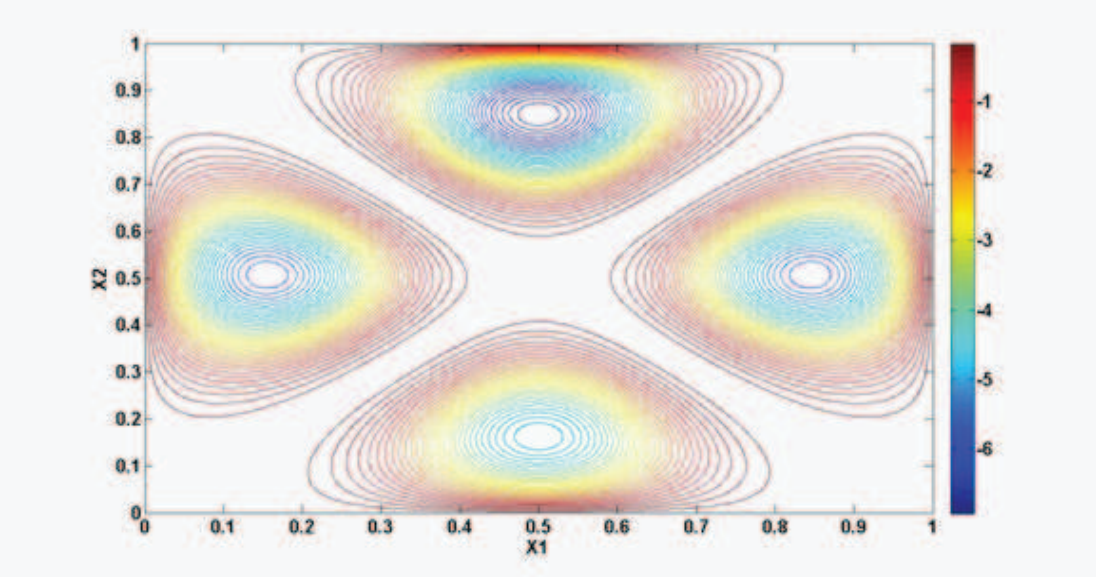


Figure 5.8: Contour plot of the two dimension test function.

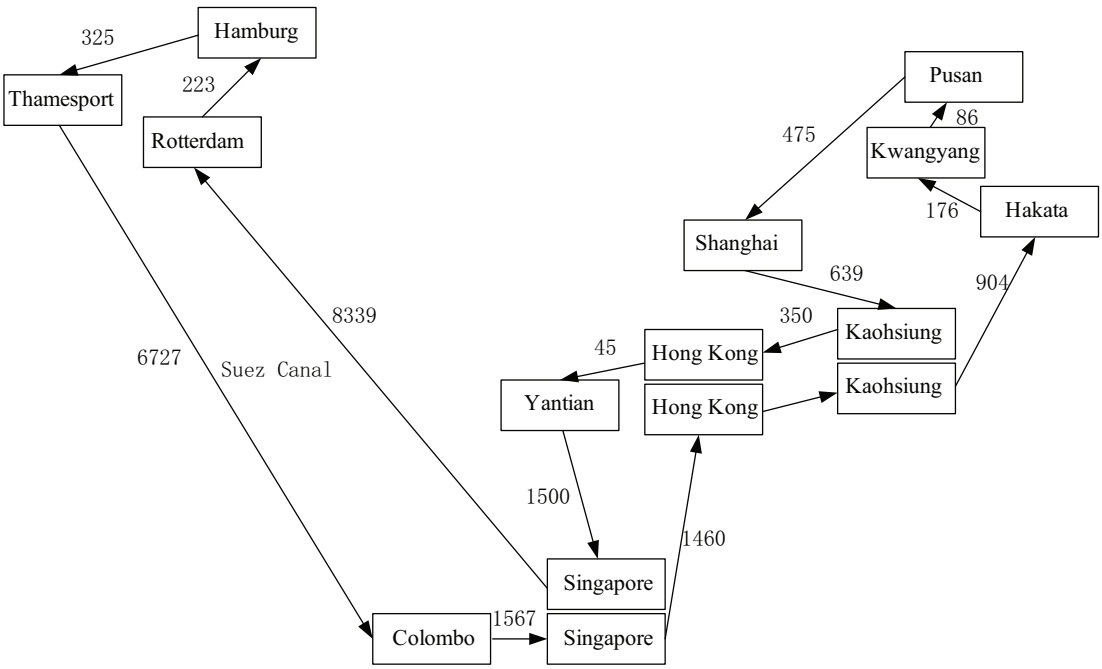


Figure 5.9: AEX service route(distances in nautical miles).

Chapter 6

BAYESIAN METAMODELING AND DESIGN APPROACH FOR STOCHASTIC SIMULATIONS

6.1 Introduction

Based on the discussion in the previous chapters, the kriging model is one of the more promising metamodels for the applications in the computer simulation as it is more adaptable than the regression based models and not as complicated and time consuming as artificial intelligence techniques. However, as indicated in Chapter 3 and Chapter 4, the performance of kriging model highly deteriorates when the noise varies, also see Yin *et al.* (2008), Li *et al.* (2010a). Yin *et al.* (2008) and Ankenman *et al.* (2010) propose the modified nugget effect model and the stochastic kriging model respectively to address the more general heteroscedastic case. Although these models enable much better modeling of stochastic systems, several issues in parameter estimation arise when the noise levels are high, as mentioned in Chapter 4.

we provide a simple quadratic test function example. Consider the test function $y = x^2 + \varepsilon(x)$, where the noise function $\varepsilon(x)$ is normally distributed with a zero mean and constant variance σ_ε^2. Suppose the response surface is unknown and

observations are obtained from 11 evenly distributed locations from -5 to 5. Two different scenarios are considered: a low variance scenario where $\sigma_\varepsilon^2 = \sigma_{\text{low}}^2 = 1$, and a high variance scenario where $\sigma_\varepsilon^2 = \sigma_{\text{high}}^2 = 10$. 50 replicates are taken at each observed location and the simulation response is estimated with a MNEK model of the form $Y(x) = Z(x) + \epsilon(x)$ where Z is a Gaussian random process that models the unknown response surface (x^2 in this case) and $\epsilon(x)$ describes the random noise inherent in the stochastic simulation. To model the unknown response surface, Z is assumed to have a constant mean β_0 and the covariance between outputs at two design points, x_i and x_j, is described by a Gaussian covariance function $\sigma_Z^2 R(Z(x_i), Z(x_j)) = \sigma_Z^2 \exp(-\phi(x_i - x_j)^2)$, where ϕ is the sensitivity parameter that controls how fast the correlation decays with the distance. With the observed data, the parameters of the model are estimated using the Maximum Likelihood Estimation (MLE) method.

Table 6.1: Empirical mean and standard deviation and quantiles of parameter estimates and the predictor's output.

estimator	scenarios	est. mean	est. std dev	est. 25th quantile	est. 75th quantile
$\hat{\beta}_0$	low variance	70.35	43.1	62.60	73.80
	high variance	69.87	44.6	55.30	81.70
$\hat{\sigma}_z^2$	low variance	2722.0	4.8425×10^3	2364.0	3139.0
	high variance	2878.0	1.2087×10^4	1628.0	3922.0
$\hat{\phi_Z}$	low variance	0.64	1.1	0.54	0.71
	high variance	0.72	1.4	0.50	0.81
$\hat{Y}(x_0)$	low variance	-0.28	0.1	-0.38	-0.19
	high variance	-0.30	0.9	-0.41	-0.14

Table 6.1 shows the empirical mean, standard deviation and quantiles of the estimated parameters and the model predictor's output at point $x_0 = 0$ in both the low variance and the high variance scenarios based on 1000 macroreplications. The variabilities of the estimated parameters in the high variance scenario

is significantly larger than in the low variance scenario. As the designs for both scenarios are fixed, this suggests that the random noise affects the parameter estimations. Furthermore, the increase in variance and variability in the parameters result in an increase of the overall prediction variability. This increase in the overall prediction variability as the simulation noise increases is also noted in Ankenman *et al.* (2010).

In Chapter 4, we considered only the sensitivity parameter ϕ_Z as unknown and estimated, and decomposed the overall prediction error into three error components, the model misspecification error, the random noise error and the parameter estimation error for stochastic simulations. As observed in Yin *et al.* (2009), the last two components increase as the variability of the random noise increases. This parameter estimation error is usually ignored when applying the traditional plug-in estimator of the predictor and the prediction error. This can lead to overconfident confidence/prediction intervals which can translate to design decisions that result in the actual system performing very differently from the expected performance in the design phase. Figure 6.1 plots the plug-in average Mean Squared Prediction Error (MSPE) of the MNEK model and the observed MSPE in the high variance scenario over 1000 macroreplications. As seen, the plug-in MSPE of the MNEK model underestimates the actual observed prediction error. This phenomenon is also highlighted and studied in den Hertog *et al.* (2005) using a bootstrapping approach for deterministic kriging.

In this chapter, we propose a Bayesian approach for kriging metamodeling for stochastic simulations. This general approach overcomes some of the problems identified in Chapter 4 by appropriately accounting for all parameter uncertainties in the model and its predictor. We discuss this metamodeling approach for certain special cases and also for the general case to fully account for all parameter uncertainties and noise levels. This modeling form can also lead to more effective ways of planning simulation experiments by balancing the need for exploring new points or regions in the design space and placing more points at observation areas to reduce the uncertainties in the metamodel form, its parameters and measurement errors. This chapter is organized as follows. In Section 2, we first describe our kriging metamodel formulation, then derive the predictive distributions of the

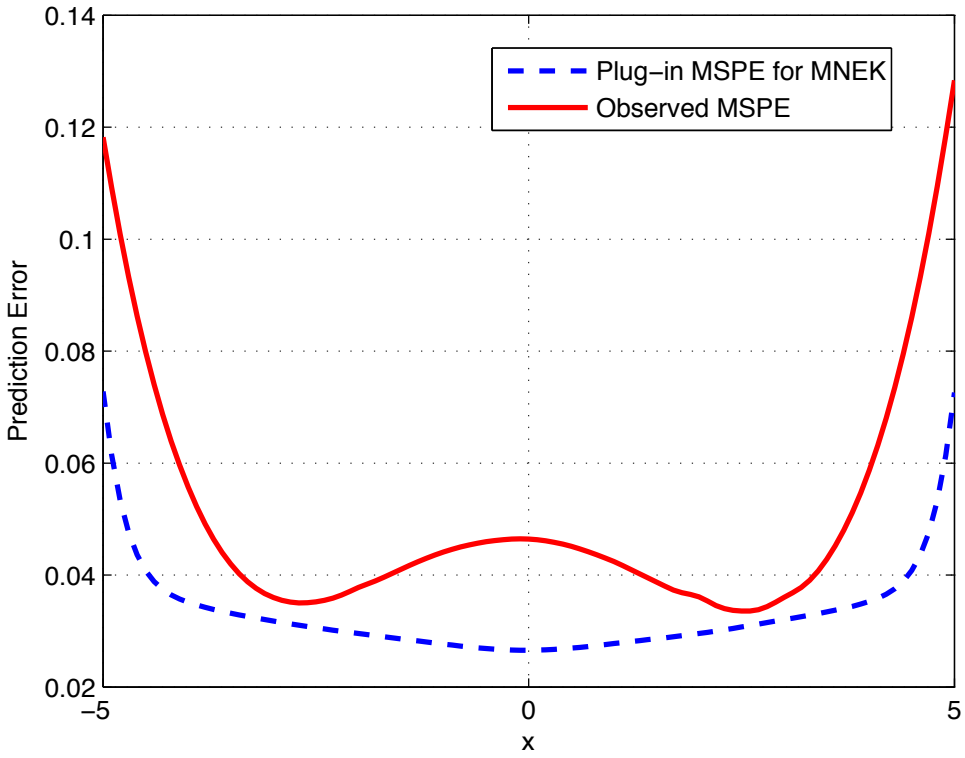


Figure 6.1: Average plug-in MSPE of MNEK and the observed MSPE for the high variance scenario.

model for specific special cases. Then, we propose a Markov Chain Monte Carlo (MCMC) approach to derive the predictive distribution for the general case.

6.2 Model Formulation

Let $Y^s(\boldsymbol{x}_i)$ be the output from the stochastic simulation at $\boldsymbol{x}_i$, where $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,p})^{\mathrm{T}}$ is a point in a p-dimensional input space. We assume that $Y^s(\boldsymbol{x}_i)$ are realizations of a random process that can be described by the model

$$Y^s(\boldsymbol{x}_i) = Z(\boldsymbol{x}_i) + \varepsilon(\boldsymbol{x}_i) \qquad i = 1, \ldots, n \tag{6.1}$$

where Z describes the mean of the process and $\varepsilon(\sim)$ describes the random noise of the process. We further assume that $Z(\boldsymbol{x})$ can be modeled as a *Gaussian*

process with mean $\mu(\boldsymbol{x}) = F^T(\boldsymbol{x})\boldsymbol{\beta}$, where $F(\boldsymbol{x})$ is a vector of q functions of $\boldsymbol{x}$, and covariance function $\sigma_Z^2 R_Z$. Here we denote this as $GP(\mu(\boldsymbol{x}), \sigma_Z^2 R_Z)$. The noise $\varepsilon(\boldsymbol{x})$ is assumed to be distributed with zero mean and covariance function $\sigma_\xi^2 K$, where K denotes the matrix of sampling correlations. Here we do not assume error variances are constant and they may depend on $\boldsymbol{x}$. With independent sampling (i.e. no CRN), K is diagonal, and Equation (6.1) reduces to the independent sampling noise model adopted by Yin *et al.* (2008). The general form of Equation (6.1) is similar to the form proposed in Ankenman *et al.* (2010).

The correlation function R_Z can have different forms, and usually depends on correlation parameter ϕ_z. A popular choice is the p-dimensional separable version of the power exponential family of correlation functions which is commonly applied in the field of computer experiments for its smooth response characteristics,

$$R_z(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)) = \prod_{k=1}^{p} \exp(-\phi_{z,k}(x_{i,k} - x_{j,k})^t). \tag{6.2}$$

The scale correlation parameter $\phi_{z,k}$ is the sensitivity parameter that controls how fast the correlation decays with distance in the kth dimension, and the parameter t controls in general the smoothness of the response. When $t = 1$, Equation (6.2) is known as the exponential correlation function, and when $t = 2$, it is known as the Gaussian correlation function. Other forms of correlation functions include the Matérn class of functions, the rational quadratic class of functions and the spline functions (more details can be found in Cressie (1993)), and the linear correlation function,

$$R_z(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)) = \begin{cases} \prod_{k=1}^{p} (1 - \phi_{z,k}|x_{i,k} - x_{j,k}|) & \text{when } \phi_{z,k}|x_{i,k} - x_{j,k}| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.3}$$

6.2.1 Modeling Uncertainty

In the model defined in Equation (6.1) and the correlation function in Equation (6.2) or Equation (6.3), the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_Z^2, \phi_Z, \sigma_\xi^2)$ are typically unknown. We write $\phi_Z = (\phi_{z,1}, \ldots, \phi_{z,p})$. From a Bayesian viewpoint, the Gaussian process $Z(\boldsymbol{x})$ defined in Equation (6.1) can be interpreted as the prior for

the unknown mean function of the stochastic simulator, with $\boldsymbol{\theta}$ as unknown parameters of the model (Currin *et al.* (1991)). To quantify the uncertainty in these parameters, a prior density $p(\boldsymbol{\beta}, \sigma_Z^2, \phi_Z, \sigma_\xi^2)$ on the parameters is assigned.

In this chapter, we assume that the parameters $\{\boldsymbol{\beta}, \sigma_Z^2, \phi_Z\}$ are independent of $\{\sigma_\xi^2\}$. We also assume that $\{\boldsymbol{\beta}, \sigma_Z^2\}$ and $\{\phi_Z\}$ are independent. Then, the prior on the parameters can be written as

$$\begin{aligned} p(\boldsymbol{\theta}) &= p(\boldsymbol{\beta}, \sigma_Z^2, \phi_Z, \sigma_\xi^2) = p(\boldsymbol{\beta}, \sigma_Z^2)p(\phi_Z)p(\sigma_\xi^2) \\ &= p(\boldsymbol{\beta}|\sigma_Z^2)p(\sigma_Z^2)p(\phi_Z)p(\sigma_\xi^2). \end{aligned} \tag{6.4}$$

This modeling approach also facilitates the accounting of any prior knowledge on the parameters. In cases with little prior information, non informative priors can be used.

6.2.2 Observed Data

For stochastic simulations, instead of directly using the observed data to fit the metamodel as in deterministic simulations, replications are typically taken at each design point. The sample mean and sample variance are obtained from the replicates and applied as inputs to fit the metamodel, see Ankenman *et al.* (2010) and Yin *et al.* (2008). In this chapter, let $Y_n = (y_1, \ldots, y_n)^{\mathrm{T}}$ denote the vector of the sample means observed at the design locations $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, $\mathbf{S}_n^2 = (s_1^2, \ldots, s_n^2)^{\mathrm{T}}$ denote the vector of the sample variances observed at the same locations, $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)^{\mathrm{T}}$ denote the vector of the number of replications taken at each design location, $\boldsymbol{X}_n$ denote the matrix of design locations and $D_n = [Y_n, \mathbf{S}_n^2]$. Typically in practice, the sample means and variances are the unbiased or MLE estimates of the means and variances at each design location.

6.2.3 Bayesian Prediction and Predictive Distribution

Based on Equation (6.1) and Equation (6.4), to predict the mean of response $Z(\boldsymbol{x}_0)$ at any $\boldsymbol{x}_0$, we first obtain the predictive distribution of $Z(\boldsymbol{x}_0)$ given D_n. The posterior predictive density at $\boldsymbol{x}_0$ can be obtained by

$$f_p(Z(\boldsymbol{x}_0)|D_n) = \int p_z(Z(\boldsymbol{x}_0)|\boldsymbol{\theta}, D_n) \times p(\boldsymbol{\theta}|D_n)d\boldsymbol{\theta}. \tag{6.5}$$

The mean $\mu_{0|D_n}(\boldsymbol{x}_0)$ of this predictive distribution is the best MSPE predictor at $\boldsymbol{x}_0$, and variance can be interpreted as the mean squared prediction error of the predictor $\mu_{0|D_n}(\boldsymbol{x}_0)$ (Santner *et al.* (2003), p. 92). The advantage of this Bayesian approach is that it accounts for the parameter uncertainties, and hence can avoid the bias introduced with point estimates as described in Zimmerman & Cressie (1992b) and Yin *et al.* (2009).

6.2.3.1 Derivation of the Predictive Distribution (*Assuming* ϕ_Z *is known*)

In this section, to ease the computational burden, we first develop a prediction procedure by assuming that the parameter ϕ_Z is known. Hence, R, the correlation matrix over all the design points and c, the correlation vector between the value of Z at the point of interest $\boldsymbol{x}_0$ and the value of Z at the design points, are known. Here, we adopt an empirical Bayes approach and discuss later how these parameters can be estimated.

With ϕ_Z known and rewriting the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_Z^2, \sigma_\xi^2)$, we have from Equation (6.5),

$$\begin{aligned} f_p(Z(\boldsymbol{x}_0)|D_n) &= \iiint p_z(Z(\boldsymbol{x}_0)|D_n, \boldsymbol{\beta}, \sigma_Z^2, \sigma_\xi^2) p(\boldsymbol{\beta}, \sigma_Z^2, \sigma_\xi^2|D_n) d\boldsymbol{\beta} d\sigma_Z^2 d\sigma_\xi^2 \qquad (6.6) \\ &= \iiint p_z(Z(\boldsymbol{x}_0)|D_n, \boldsymbol{\beta}, \sigma_Z^2, \sigma_\xi^2) p(\boldsymbol{\beta}|\sigma_Z^2, D_n) p(\sigma_Z^2, \sigma_\xi^2|D_n) d\boldsymbol{\beta} d\sigma_Z^2 d\sigma_\xi^2 \end{aligned}$$

and we allow the error variances to be dependent on $\boldsymbol{x}_i$.

6.2.3.2 Modeling of σ_ξ^2

In general, for most stochastic simulation problems, the variance of the noise $\mathrm{Var}(\varepsilon) = \sigma_\xi^2$ is unknown, and in our model treated as a random variable. Moreover, σ_ξ^2 may be non constant and dependent on $\boldsymbol{x}_i$. Treating σ_ξ^2 as an unknown function (which can depend on $\boldsymbol{x}$), we model $\log(\sigma_\xi^2)$ (which we denote as V) as a Gaussian process, $V \sim GP(\mu_V, \sigma_V^2 R_V)$, where R_V can take similar forms as R_z (e.g. Equation (6.2) or Equation (6.3)). This Gaussian process model can be interpreted as a prior for the unknown function of the natural logarithm of the simulation variance. Parameters of this Gaussian process prior can be estimated

with the observed data $\boldsymbol{V}_n = [\log(s_1^2), \log(s_2^2), \ldots, \log(s_n^2)]$, to yield the posterior process for $\log(\sigma_\xi^2)$. The distribution for σ_ξ^2 can then be obtained by a straightforward transformation of V. The approach taken in Ankenman *et al.* (2010) can be similarly interpreted, although in this chapter, we place a Gaussian process prior on $\log(\sigma_\xi^2)$ instead of σ_ξ^2. This natural log transformation has the nice properties of approximating normality, stabilizing variance and ensuring inverse transformation back to the positive scale. Alternative transformations are also possible (see Bekki *et al.* (2009)).

Describing $\sigma_\xi^2(\cdot)$ in this manner enables the error covariances to be modeled as functions of locations and can also provide estimation of the noise at unobserved points by "borrowing" information on variability from other nearby points. This formalizes the ideas in den Hertog *et al.* (2005) and Yin *et al.* (2008) for estimating variances at unobserved points where ad hoc interpolation and bootstrapping techniques were taken. Although different in functional forms, similar ideas of modeling the variances to assist in better estimation and prediction of the mean responses are taken in Cheng & Kleijnen (1999a).

A simple two point problem We use a simplified two point problem with a one dimensional input space to illustrate the approach analytically and provide some insights.

Suppose two sample means y_1 and y_2, and sample variances s_1^2 and s_2^2 are observed at two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ respectively. We assume the noise process ε has homogeneous variance, $\sigma_\xi^2(\boldsymbol{x}_1) = \sigma_\xi^2(\boldsymbol{x}_2) = \sigma_\xi^2$, and the number of replications used at both design locations is r_0. We further assume that r_0 is sufficiently large that the observed sample variances (computed using MLE method) is approximately $s_1^2 = s_2^2 = s_0^2$. The distance between these two points is d_0, and the prediction point $\boldsymbol{x}_0$ is located at the center of the two points. We assume a linear correlation between $Z(\boldsymbol{x}_1)$ and $Z(\boldsymbol{x}_2)$ with the linear correlation function given in Equation (6.3), here simplified to $\rho_z = 1 - \phi_Z |d_{ij}|$, where $d_{ij} = \boldsymbol{x}_i - \boldsymbol{x}_j$. Hence, the correlation for this two point problem is given as $\rho_0 = 1 - \phi_Z |d_0|$. We also assume a constant mean function (i.e. $F^T(\boldsymbol{x}_0)\boldsymbol{\beta} = \beta_0$) for the mean process, and that the observations obtained at the design locations are independent. To facilitate the integration, we introduce a new parameter τ defined as $\tau = \sigma_\xi^2 / \sigma_Z^2$, indicating

loosely the ratio of the error noise to the signal function variation. For computational simplicity, we adopt "Jeffreys priors" (see Jeffreys (1961) p. 179) for β_0, σ_Z^2 and τ with the following forms: $p(\beta_0) = 1, p(\sigma_Z^2) = 1/\sigma_Z^2, p(\tau) = 1/(2(1+\tau)^2)$, for $\beta_0 \in \Re, \sigma_Z^2 > 0$ and $\tau > 0$. The priors for β_0 and σ_Z^2 are similar to the non informative priors proposed in Santner *et al.* (2003) (p. 94) for Gaussian random function models of similar forms.

Then the covariance matrix of the observations $\mathbf{\Sigma}$, covariance vector $v(\boldsymbol{x}_0)$, observation vector Y_n and $\mathbf{S}_n^2$, and the regressors for the observation points and prediction point are given respectively as

$$\mathbf{\Sigma} = \sigma_Z^2 R = \begin{pmatrix} \sigma_Z^2 & \sigma_Z^2\rho_0 \\ \sigma_Z^2\rho_0 & \sigma_Z^2 \end{pmatrix} + \begin{pmatrix} \sigma_Z^2\frac{\tau}{r_0} & 0 \\ 0 & \sigma_Z^2\frac{\tau}{r_0} \end{pmatrix}, \; v^{\mathrm{T}} = \sigma_Z^2 c^{\mathrm{T}} = \sigma_Z^2 \left(\frac{1+\rho_0}{2}, \frac{1+\rho_0}{2} \right),$$
$$\rho_0 = 1 - \phi_z|d_0| \;, \; Y_n^{\mathrm{T}} = (y_1, y_2) \;, \; \mathbf{S}_n^{2\mathrm{T}} = (s_0^2, s_0^2) \;, \; F^T(\boldsymbol{x}) = (1,1) \;, \; F(\boldsymbol{x}_0) = 1.$$

The predictive distribution at $\boldsymbol{x}_0$ is then given as (see proof in Appendix A)

$$f_p(Z(\boldsymbol{x}_0)|D_n) \propto \left(1 + \frac{1}{2} \frac{(Z(\boldsymbol{x}_0) - \mu_{|D_n}(\boldsymbol{x}_0))^2}{\sigma_{|D_n}^2(\boldsymbol{x}_0)} \right),$$

where $\mu_{|D_n}(\boldsymbol{x}_0) = \frac{(y_1+y_2)}{2}$ and

$$\begin{aligned} \sigma_{|D_n}^2(\boldsymbol{x}_0) \;\; = \;\; & \frac{(1-\rho_0)(y_1-y_2)^2}{2} + \frac{s_0^2}{r_0} + \{(s_0^2(1-\rho_0) + (y_1-y_2)^2\rho_0[\ln(r_0(1-\rho_0)) \\ & (1-\rho_0) + \rho_0])/(2r_0[\ln(r_0(1-\rho_0))(1-\rho_0) + \rho_0])\}. \end{aligned} \tag{6.7}$$

Under these modeling assumptions, the predictive distribution at $\boldsymbol{x}_0$ is a non-central t-distribution with mean $\mu_{|D_n}(\boldsymbol{x}_0)$ and variance $\sigma_{|D_n}^2(\boldsymbol{x}_0)$. With the distribution results, other characteristics of the predictor can be obtained, and the 100(1-α)% prediction interval is given as $\mu_{|D_n}(\boldsymbol{x}_0) \pm \sigma_{|D_n}^2(\boldsymbol{x}_0) \cdot t_{r_0,\alpha/2}$.

Comparing these results with the predictive results of the MNEK model, we see that this predictive mean is equivalent to the predictor obtained from the MNEK model. The predictive variances are more indicative of the effect of noise on parameter estimation and the overall prediction. Comparing Equation (6.7) with the MSPE of the predictor of the MNEK model $\left(\frac{r_0(1-\rho_0)(y_1-y_2)^2+2s_0^2}{2r_0}\right)$, we see that the last two terms of the predictive variance (in the curly brackets of Equation (6.7)) is the additional uncertainty not accounted for in the MNEK model.

From Equation (6.7), we see that not only does the predictive variance increase as the noise level s_0^2 increases, but that this additional variance component increases with the noise level too. This indicates that the MSPE increasingly underestimates the predictive variance as the noise level increases. From Equation (6.7), we also see that the predictive variance will decrease as the number of replications r_0 increase, and that the additional variance component will also decrease as r_0 increases. This indicates that the effect of r_0 not only reduces the direct impact of s_0^2, but also reduces the additional variability contributed from parameter estimation. This additional reduction shows that the underestimation of the MSPE of the MNEK predictor can reduce as the number of replications increases. Furthermore, from Equation (6.7), we observe that the effect of ρ_0 will depend on the balance between $(y_1 - y_2)^2$ (variability in estimating the mean response surface) and s_0^2 (inherent noise in the simulation).

6.2.3.3 A further simplification of Equation (6.6)

Assume $\tau(\boldsymbol{x}) = \sigma_\xi^2(\boldsymbol{x})/\sigma_Z^2$ is known (we discuss the estimation of τ in the next subsection) and the following hierarchical proper prior setup

$$\boldsymbol{\beta}|\sigma_Z^2 \sim N_p(\mathbf{w}_0, \sigma_Z^2\mathbf{Q}_0) \ , \quad \sigma_Z^2 \sim IG(\alpha_z, \gamma_z) \ , \quad \phi_{z,k} \sim G(a_z, b_z) \text{ for } k = 1, 2, \ldots, p \tag{6.8}$$

where $N_p(\mathbf{u}, \mathbf{H})$ denotes a p-dimensional normal with mean $\mathbf{u}$ and covariance matrix $\mathbf{H}$, $IG(\alpha, \gamma)$ denotes a Inverse Gamma distribution with density function,

$$p(s) = \frac{\gamma^\alpha}{\Gamma(\alpha)} s^{-(\alpha+1)} e^{-(\frac{\gamma}{s})}, s > 0, \alpha, \gamma > 0,$$

and $G(a, b)$ denotes the gamma distribution with density function

$$p(s) = \frac{b^a}{\Gamma(a)} s^{a-1} e^{-bs}, s > 0, a, b > 0.$$

This selection of prior distributions follows the conjugate prior for the Gaussian Random Process in Gelman *et al.* (2003), and similar prior settings can be found in Santner *et al.* (2003) (page 94). Qian & Wu (2008) suggest values for hyperparameters $\mathbf{w}_0, \mathbf{Q}_0, \alpha_z, \gamma_z, a_z, b_z$ that reflect "vague" knowledge or flat priors on the

parameters. When prior expert knowledge is available, Oakley (2002) provides detailed methods for eliciting these hyperparameters from observable quantities.

Further assuming independent sampling at each design point, we obtain the following proposition.

Proposition 6.2.1 *If ϕ_Z and τ are known, then the predictive distribution at $\boldsymbol{x}_0$ is a non-central t distribution*

$$Z(\boldsymbol{x}_0)|D_n \sim T_1(\alpha_{Z|D_n}, \mu_{|D_n}(\boldsymbol{x}_0), \sigma^2_{|D_n}(\boldsymbol{x}_0))$$

where

$$\alpha_{Z|D_n} = n + 2\alpha_Z, \;\; \mu_{|D_n}(\boldsymbol{x}_0) = F(\boldsymbol{x}_0)\mathbf{M}^{-1}\boldsymbol{\lambda} + c^{\mathrm{T}}(\boldsymbol{x}_0)R^{-1}(Y_n - F(\boldsymbol{x})\mathbf{M}^{-1}\boldsymbol{\lambda}),$$

and

$$\begin{aligned}\sigma^2_{|D_n}(\boldsymbol{x}_0) \;\; = \;\; & \frac{2\gamma_Z + Y_n^{\mathrm{T}}R^{-1}Y_n + \mathbf{w}_0^{\mathrm{T}}\mathbf{Q}_0^{-1}\mathbf{w}_0 - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{M}^{-1}\boldsymbol{\lambda}}{2\alpha_{Z|D_n}} \\ & *(1 - c^{\mathrm{T}}(\boldsymbol{x}_0)R^{-1}c(\boldsymbol{x}_0) + \mathbf{k}\mathbf{M}^{-1}\mathbf{k}^{\mathrm{T}}),\end{aligned}$$

with

$$\boldsymbol{\lambda} = F^T R^{-1}Y_n + \mathbf{Q}_0^{-1}\mathbf{w}_0, \;\; \mathbf{M} = F^T R^{-1}F + \mathbf{Q}_0^{-1}, \;\; \mathbf{k} = c^{\mathrm{T}}(\boldsymbol{x}_0)R^{-1}F - F(\boldsymbol{x}_0),$$

$R = R_Z + R_\xi$ *and* $R_\xi = diag\,[\tau(\boldsymbol{x}_1)/r_1 \ldots \tau(\boldsymbol{x}_n)/r_n]$.

See Appendix E.

Based on the above, a 100(1-α)% prediction interval for any point $\boldsymbol{x}_0$ is given as $\mu_{|D_n}(\boldsymbol{x}_0) \pm \sigma^2_{|D_n}(\boldsymbol{x}_0) \cdot t_{\alpha_{Z|D_n}/2,\alpha/2}$, where $t_{\alpha_{Z|D_n}/2,\alpha/2}$ is a univariate non central t distribution with $\alpha_{Z|D_n}/2$ degrees of freedom. In this special case, $D_n = [\boldsymbol{X}_n, Y_n, \boldsymbol{r}]$ as τ is assumed known.

Estimation of parameters ϕ_Z and τ The results above are based on the assumption that the parameters ϕ_Z and τ are known. We adopt an empirical Bayes approach here by estimating these parameters for computational convenience. Here we propose to estimate these parameters using the MLE method. This is similar to the approach taken by Wang *et al.* (2009). In Section 6.2.3.4, we consider a fully Bayesian approach which explicitly incorporates the uncertainties

of these parameters in the modeling framework and propose a MCMC approach to obtain the predictive distribution.

Applying the MLE method, τ and ϕ_Z are the parameter values that maximize the marginal likelihood $p_y(Y_n|\tau, \phi_Z)$. Hence,

$$\{\hat{\tau}, \hat{\phi_Z}\} = \arg\max_{\tau,\phi_Z} p_y(Y_n|\tau, \phi_Z)$$

With independent sampling, $p_y(Y_n|\tau, \phi_Z)$ can be derived as

$$\begin{aligned} p_y(Y_n|\tau, \phi_Z) &= \frac{1}{n}\log(\det(R(\tau, \phi_Z))) + \log\left[\frac{1}{n}\left(Y_n - F\left(F^T R(\tau, \phi_Z)^{-1} F\right)^{-1}\right.\right. \\ &\quad \left.\left.\left(F^T R(\tau, \phi_Z)^{-1} Y_n\right)^{\mathrm{T}} R(\tau, \phi_Z)^{-1}\left(Y_n - F\frac{F^T R(\tau, \phi_Z)^{-1} Y_n}{F^T R(\tau, \phi_Z)^{-1} F}\right)\right)\right] \end{aligned}$$

where (τ, ϕ_Z) is included in the notation for the correlation matrix R here to indicate that R is a function of the variance ratio τ and the sensitivity parameter ϕ_Z. For a given set of design points and observations, this can be solved using standard optimization functions in MATLAB. Multiple starting values can be applied in the optimization to overcome a potential problem of multiple local optimums.

Alternatively, as τ represents the ratio of the noise variance to the signal variation, the sample variance of the noise process and the signal process can be used to first estimate the ratio τ. This can be estimated at each location $\boldsymbol{x}_i$ by

$$\hat{\tau}(\boldsymbol{x}_i) = \frac{s_i^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2/(n-1)} \tag{6.9}$$

where s_i^2 represents the sample variance observed at location $\boldsymbol{x}_i$, y_j represents the sample mean observed at location $\boldsymbol{x}_j$, and $\bar{y}$ is the grand mean of all the observed sample means at all locations $\boldsymbol{x}_1$ to $\boldsymbol{x}_n$. With the $\hat{\tau}$ estimated from the observed sample variance, estimating ϕ_Z simplifies and reduces to maximizing marginal likelihood $p_y(Y_n|\phi_Z, \hat{\tau}(\boldsymbol{x}_i))$, i.e.

$$\{\hat{\phi_Z}\} = \arg\max_{\phi_Z} p_y(Y_n|\phi_Z, \hat{\tau}(\boldsymbol{x}_i))$$

A different approach to using the above MLE method is to first estimate the posterior distribution of the parameters ϕ_Z and τ, and then using the posterior

mean or mode of these distributions to estimate ϕ_Z, and τ. This approach avoids the potential instability and local maxima issues of the MLE method. A similar approach to estimate these parameters is taken by Qian & Wu (2008).

6.2.3.4 A General Approach to Deriving the Predictive Distribution (when all parameters are unknown)

In general, the integration in Equation (6.5) can be done numerically. Assuming the prior setup in Equation (6.8) and a Gaussian process prior on $\log(\sigma^2_\xi(\boldsymbol{x}_0))$, and the full conditionals $p(\phi_Z, \phi_V|Y_n, \mathbf{S}^2_n, \sigma^2_Z, \sigma^2_V)$, $p(\sigma^2_Z|Y_n, \mathbf{S}^2_n, \phi_Z)$, $p(\sigma^2_V|\mathbf{S}^2_n, \phi_V)$ and $p(\sigma^2_\xi(\boldsymbol{x}_0)|\mathbf{S}^2_n, \phi_V)$ (in Appendix C), an MCMC approach, specifically, the Gibbs sampler, can be applied in the following manner. For convenience of notation here, we omit the dependence of the conditionals on $\boldsymbol{X}_n$ and $\boldsymbol{r}$.

1 Set an initial starting value for $(\sigma^2_Z)^0, (\sigma^2_V)^0$

2 Sample $(\phi_Z, \phi_V)^i$ from $p(\phi_Z|Y_n, (\sigma^2_Z)^{i-1}) \cdot p(\phi_V|\mathbf{S}^2_n, (\sigma^2_V)^{i-1})$

3 Sample $(\sigma^2_Z)^i$ from $p(\sigma^2_Z|Y_n, (\phi_Z)^i)$

4 Sample $(\sigma^2_V)^i$ from $p(\sigma^2_V|\mathbf{S}^2_n, (\phi_V)^i)$

5 Sample $(\sigma^2_\xi(\boldsymbol{x}_0))^i$ from $p(\sigma^2_\xi(\boldsymbol{x}_0)|\mathbf{S}^2_n, (\phi_V)^i)$

6 Sample $Z(\boldsymbol{x}_0)^i$ from $f_p(Z(\boldsymbol{x}_0)|Y_n, (\sigma^2_\xi(\boldsymbol{x}_0))^i, (\sigma^2_Z)^i, (\phi_Z)^i)$

7 Repeat steps 2 to 6 until the chain has converged or a stopping criterion is met

8 Approximate $\mu_{|D_n}(\boldsymbol{x}_0)$ by $\hat{\mu}_{|D_n} = \frac{\sum_{i=k-m}^{k} Z(\boldsymbol{x}_0)^i}{(k-m)}$ and $\sigma^2_{|D_n}(\boldsymbol{x}_0)$ by $\hat{\sigma}^2_{|D_n} = \frac{\sum_{i=k-m}^{k}(Z(\boldsymbol{x}_0)^i - \hat{\mu}_{|n})^2}{(k-m)-1}$

where k is the length of the chain and m is the initial burn in samples. The prediction interval for the predictor $\hat{\mu}_{|n}$ can be obtained from the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of the $Z(\boldsymbol{x}_0)^i$ samples. For assessing the convergence in step 7, related discussions can be found in Cowles & Carlin (1996) and Gelman *et al.* (2003). These include practical stopping criteria based on sample statistics collected from the chain.

6.3 Numerical Examples

To illustrate this general Bayesian metamodeling approach with the MCMC algorithm, we conduct numerical examples on the M/M/1 queueing system and a quadratic function. We consider both the Bayesian predictive model with the simplification when the correlation function and τ are known (BKS) and the Bayesian predictive model with the MCMC sampling (BKMCMC). In this section, both of the models are tested over two different examples. In the analysis of these examples, we use general "vague" hyperparameter setting for the priors similar to those adopted in Qian & Wu (2008). Specifically, we chose a "vague" prior of $IG(2,1)$ for σ_Z^2, σ_V^2, the "location-flat" prior $N(\mathbf{0}, \sigma_Z^2), N(\mathbf{0}, \sigma_V^2)$ for the $\boldsymbol{\beta}, \boldsymbol{\beta}_V$, and $G(2,1)$ for ϕ_Z, ϕ_V. For the BKS model, τ is estimated using Equation (6.9).

In order to study the effects of the noise level on the parameter estimates and predictions, we compare our Bayesian predictive models with that of the Modified Nugget Effect Kriging (MNEK) model (Yin *et al.* (2008)). The parameters of the MNEK model are estimated using the MLE method. Since the mathematical form of the MNEK and the stochastic kriging model (Ankenman *et al.* (2010)) are equivalent, we will make the comparison only with MNEK.

6.3.1 The Simple Quadratic Function

Consider the simple quadratic function in Section 6.1.

$$y = x^2 + \varepsilon(x)$$

where the $\varepsilon(x)$ is the noise function with a zero mean and a constant variance σ_ε^2 when $x \in [-5, 5]$. 11 observations are evenly distributed from -5 to 5. We consider two scenarios (a high variance and a low variance scenario) in this example. For the high variance scenario the variance of the random noise function is $\sigma_{\text{high}}^2 = 1000$, and for the low variance scenario, $\sigma_{\text{low}}^2 = 10$. The number of replications taken at each observed point is 100. All the three models are tested over the 1000 prediction points evenly distributed from -5 to 5. Figure 6.2 gives the predictive mean for all the three models. This figure highlights an interesting effect of noise on the kriging model's behavior. Observing first the models when

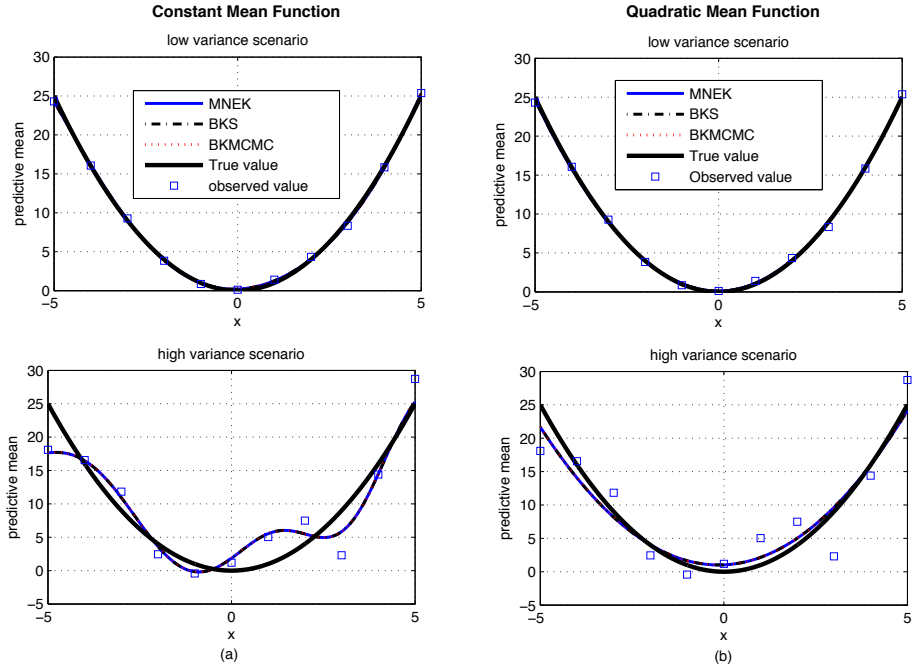


Figure 6.2: Predictive mean given by MNEK, BKS and BKMCMC for the simple quadratic function example for the low and high variance scenarios (with (a) constant and (b) quadratic mean functions).

assuming a constant mean function (Figure 6.2(a)), we see that the prediction means for all the three models are very close to the true mean in the low variance scenario. However, for the high variance scenario, the predictive mean given by all the three models oscillates around the true function. The kriging model is a spatial correlation model consisting of a mean function and a spatial correlated weighting function. The mean function is usually assumed to be a simple polynomial regression model. In cases where no prior information of the true function is available, the constant mean function is typically chosen. With the spatial correlation feature, the kriging model can often still provide satisfactory prediction results for a complicated function with a constant mean function. However, in the presence of noise, the spatial correlation parameters can be poorly estimated and the relationships erroneously weakened. As a result, especially when noise is high, the predictive mean reverts to the assumed mean function (the constant mean function in this case). This phenomenon can be explained by observing

the predictive mean in Proposition 6.2.1. When noise is large, τ increases and R gets large, the predictive mean reverts to the prior assumed mean function. This similar observation can be made for the MNEK model and the stochastic kriging model when R gets large (see Equation 6 in Ankenman *et al.* (2010)). Hence, the prediction results can be far away from the true value when the assumed mean function for the kriging model is very different from the true function, as observed in the high variance scenario with constant mean function in Figure 6.2(a). This similar phenomenon is also observed when the mean function is assumed to be a first order polynomial. This can be rectified by assigning a more appropriate underlying mean function for the mean process model (see Figure 6.2(b) when a second-order polynomial mean function is assumed).

The predictive variance of the three models with both constant and quadratic underlying mean function are shown in Figure 6.3. For the constant mean func-

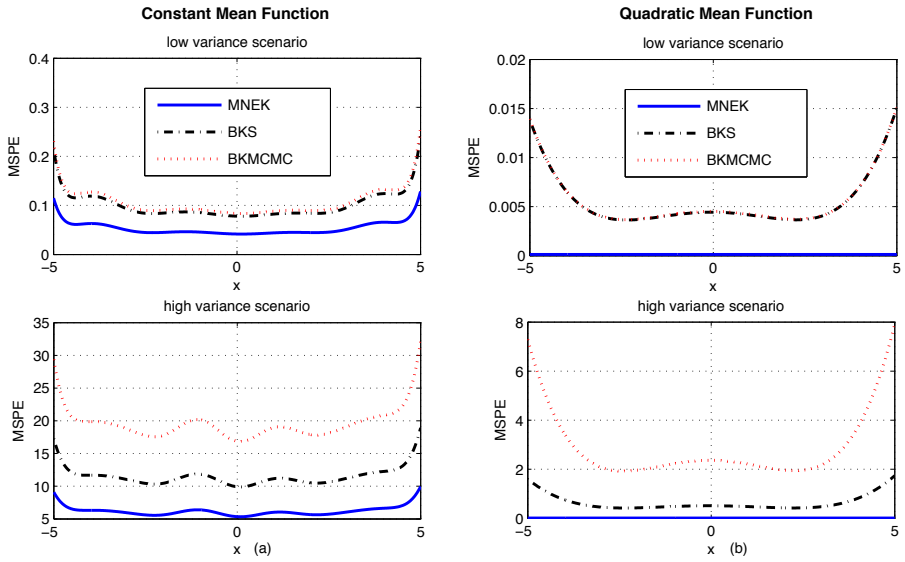


Figure 6.3: Predictive variance given by MNEK, BKS and BKMCMC for the simple quadratic function example (with (a) constant and (b) quadratic mean functions).

tion case, we see that the predictive variances provided by BKS and BKMCMC are higher than predictive variance provided by MNEK. This phenomenon is expected as the BKS and BKMCMC account for the parameter uncertainties in the

prediction error. Furthermore, the BKMCMC consider all the parameter uncertainties in the model. These parameter estimation uncertainties increase as the variability of the observed data increase. As a result, it has a higher predictive variance than the BKS in the high variance scenario. Figure 6.3 also illustrates the impact of the underlying mean function. The predictive variances for the quadratic mean function for all the three models are reduced substantially. This highlights the importance of selecting an appropriate underlying mean function especially when noise is high and prior information of the target process is available.

In order to evaluate the estimated variability and prediction intervals of the three models, we test the predictive coverage for all three models. Table 6.2 provides the average coverage of 90% prediction intervals for all three models (with quadratic mean function) based on 1000 macroreplications over 1000 test points for both scenarios.

Table 6.2: Average empirical coverage of the 90% predictive interval for MNEK, BKS and BKMCMC.

Model	LowVar ($r_0 = 10$)	HighVar ($r_0 = 10$)	LowVar ($r_0 = 100$)	HighVar ($r_0 = 100$)
MNEK	0.80	0.69	0.85	0.85
BKS	0.82	0.74	0.86	0.86
BKMCMC	0.83	0.77	0.87	0.87

As expected, when the number of replications (r_0) taken increases, the coverage for all models improve. Overall, BKMCMC provides coverage closest to the nominal coverage and BKS is a close approximation. MNEK has the lowest coverage. Paired t-tests results at each test point confirm that the coverage for MNEK with BKS and BKMCMC are statistically different (at $\alpha = .05$) while BKS and BKMCMC are similar in the low variance scenario, and the three models are statistically different in the high variance scenario when $r_0 = 10$. These results suggest that the BKS is a reasonable option in cases where variability is low as it requires less computational effort than the BKMCMC and has very

similar prediction results to BKMCMC. However, in the case where variability is high, the BKMCMC is preferred as its predictive mean and variance are closer to that observed from the simulations.

6.3.2 The M/M/1 System

The M/M/1 queue system is a typical stochastic system which is widely studied in the literature. Since the expected waiting time in queue for the M/M/1 system has a closed form, it is easy to compare the model's performance with the true performance. In this stochastic simulation system, we treat traffic rate as the input and the expected waiting time in queue as the output. With the system service rate fixed at 1 and varying the arrival rate x, the steady state expected waiting time in queue can be expressed as:

$$\mu(x) = \frac{x}{1-x}$$

As the variability in low traffic rate regions is low, the additional variability in the parameter estimates is small and not significant. For this reason, we focus in the higher traffic regions. 8 observation points are evenly located at $x = 0.50, 0.55, \ldots, 0.80, 0.85$, where the traffic rate x is considered to be the input of the simulation model. For each observation point, 100 replications were conducted, each with 100,000 arrivals, and the sample mean and sample variance taken as the simulation outputs. Based on the 8 sets of input-output combinations, we build the MNEK, BKS and BKMCMC with a constant underlying mean function (assuming little is known about the response surface). All the models are then tested over 100 evenly distributed prediction points in the sample space. For the MNEK, the predictor's output and the MSPE of the predictor are used as the evaluation criterion. For the BKS and BKMCMC, the predictive mean and predictive variance are used. The predictive mean and predictive variance of the BKMCMC are approximated over all the MCMC samples.

For comparison, we use the mean squared error (MSE) of both the predictive mean and variance to the 100 test points as the evaluation criterion. The predictive mean and variance of all three models are compared with the steady state

function and empirical variance for each test point, respectively. The average MSE is calculated over 100 test points as

$$\text{MSE}_{\text{mean}} = \frac{\sum_{i=1}^{100} (\mu_{x_i} - Z(x_i))^2}{100}, \quad \text{MSE}_{\text{var}} = \frac{\sum_{i=1}^{100} \left(\sigma_{x_i}^2 - \sigma_{empr}^2\right)^2}{100},$$

where σ_{empr}^2 is the empirical variance of the predictor, computed from 1000 independent predictor's outputs at each point. The overall average $\overline{\text{MSE}}_{\text{mean}}$ and $\overline{\text{MSE}}_{\text{var}}$ is then computed over 1000 macro replicates of the above.

Table 6.3: $\overline{\text{MSE}}$ for MNEK, BKS and BKMCMC for the M/M/1 example.

Criterion	MNEK	BKS	BKMCMC
$\overline{\text{MSE}}_{\text{mean}}$	9.3×10^{-4}	9.3×10^{-4}	9.1×10^{-4}
$\overline{\text{MSE}}_{\text{var}}$	6.8×10^{-5}	5.6×10^{-5}	1.1×10^{-5}

From Table 6.3, we see that all three models provide satisfactory results in terms of predictive mean and predictive variance. Although the response studied is different from the M/M/1 response studied in Ankenman *et al.* (2010) (average waiting time versus average number in queue), the general finding that the stochastic models from kriging fits relatively well for both the predictive mean and variance is similar. Among all the three models, BKMCMC yields the best prediction results with lowest prediction error.

However, for this M/M/1 example, the performances of all the three models deteriorate as the traffic rate x increases. In order to better illustrate how the increasing variability affects all the three models' performances, Table 6.4 lists the empirical coverage for 90% prediction intervals developed based on $r_0 = 10$ and $r_0 = 100$ initial replications at various traffic rate x levels, over 1000 macroreplicates.

Table 6.4 indicates the deterioration of the performances of all three models as the traffic rate increases. The coverage improves when the number of initial replications at each design point increases. In the high traffic regions (> 0.8), the prediction intervals constructed from the MNEK model results in very poor coverage (e.g. 51% versus 90% nominal) as the model parameters are likely to

Table 6.4: Empirical coverage of the 90% predictive interval for MNEK, BKS and BKMCMC at given points.

Location	MNEK	BKS	BKMCMC	MNEK	BKS	BKMCMC
		($r_0 = 10$)			($r_0 = 100$)	
$x = 0.55$	0.87	0.89	0.90	0.90	0.91	0.92
$x = 0.65$	0.83	0.84	0.87	0.83	0.84	0.86
$x = 0.75$	0.82	0.85	0.86	0.84	0.86	0.88
$x = 0.85$	0.51	0.56	0.74	0.65	0.72	0.80

be poorly estimated in these high variability regions and the plug-in MSPE for the MNEK model does not adequately account for this. The coverage for BKS is higher, and the coverage for BKMCMC is closest to the nominal 90%. When the underlying mean function is assumed known (using $\frac{x}{1-x}$), the $\overline{\text{MSE}}_{\text{mean}}$ for the MNEK model decreases by an order of 10, while it decreases by an order of 100 for both the BKS and BKMCMC models. The $\overline{\text{MSE}}_{\text{var}}$ decreases by an order of 100 for three models. The coverage for all models improve, increasing by almost 15% in the high traffic regions.

Overall, the Bayesian metamodeling approaches (both the BKS and BKMCMC) better estimate the true variability and provide better coverage than the MNEK for stochastic simulations, especially for cases with high variability. Therefore, properly accounting for the uncertainties in stochastic simulations can more accurately reflect the uncertainty in the metamodel and its predictor and better assist in making informed decisions. Furthermore, the BKS can be used as the simplification of the fully Bayesian BKMCMC approach in cases with low variability. In addition, as noted in both examples, selecting an appropriate mean function is crucial for all kriging model forms especially when noise is high in order to reduce the error caused by a misspecified underlying mean function.

6.4 Sequential Experimental Design Approach

The development of the kriging model in Section 6.2 raises naturally the initial experiment design problem and the issue of allocating limited computing budget. In this section, we address this problem via a two stage design framework that aims to balance the allocation of new design points and replication numbers in order to improve the predictive ability of the metamodel by reducing the overall MSPE. These considerations are useful, especially in cases where the noise levels are high at existing design points and more replications at these locations are required to drive this down. Other sequential design approaches can also be developed for other specific purposes like optimization. As the focus in this chapter is on the development and improvement of the model for prediction, we do not further discuss these alternatives.

As seen from Proposition 6.2.1, the MSPE is related to the number of replications at each design point and also to the location of the design points. A good design hence should optimally balance between these two in order to reduce the overall MSPE. In this chapter, we use the integrated mean squared prediction error (IMSPE) as a measure of the overall MSPE. Below, we first outline our two stage framework and describe our design criterion. We further decompose our criterion in terms of the design point chosen for the case when $\boldsymbol{\phi}$ and τ are known to provide some design insights. We then propose two possible approaches to handle the second stage design and provide further insights on the allocation of replications and new design points.

6.4.1 The two stage design framework

Stage 1 Use an initial space filling or uniform design. Once the data is obtained, build the initial kriging model.

Stage 2 Based on the initial model, with an appropriate design criterion, determine a follow-up experiment to augment the initial experiment and refit the kriging model with the entire data set.

Stage 1 is the typical approach taken in the general design of experiments when no previous data is available. For stage 2, in order to improve the predictive

capability of the model, we focus on reducing the uncertainties in the prediction (i.e. reduce the predictive variance or MSPE) over the input space. This is a general framework with two stages. For a practical design problem with computing budget constraints, this framework can be extended into a sequential approach by recursively applying stage 2.

6.4.2 A follow-up design criterion

Suppose initial observations $Y_{n_1} = (y_1, \ldots, y_{n_1})^{\mathrm{T}}$, $\mathbf{S}^2_{n_1} = (s_1^2, \ldots, s_{n_1}^2)^{\mathrm{T}}$ at design locations $\boldsymbol{X}_{n_1} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{n_1}]$ with $\boldsymbol{r} = (r_1, r_2, \ldots, r_{n_1})^{\mathrm{T}}$ replications were obtained from the initial experiment. Let $D_{n_1} = \left[Y_{n_1}, \mathbf{S}^2_{n_1}\right]$. The IMSPE can be expressed as

$$\text{IMSPE} = \int_{\boldsymbol{x}_0} \sigma^2_{|D_{n_1}}(\boldsymbol{x}_0) d\boldsymbol{x}_0$$

where $\sigma^2_{|D_{n_1}}(\boldsymbol{x}_0)$ is the posterior predictive variance of process Z at point $\boldsymbol{x}_0$ after observations are taken at n_1 initial design locations $\boldsymbol{X}_{n_1}$. As discussed in the previous sections, the MSPE (and hence IMSPE) can be affected by both the number of the replications and the design points chosen.

In the second stage, we propose this measure as a criterion to plan and select appropriate experimental designs. Suppose the initial design e_1 contains n_1 design points $\boldsymbol{X}_{n_1} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{n_1}]^{\mathrm{T}}$, each with r_0 replicates. At each design point $\boldsymbol{x}_i$, the output (y_i, s_i^2) is observed, where y_i is the sample mean of the r_0 replicates $y_{i1}, y_{i2}, \ldots, y_{ir_0}$ and s_i^2 is the sample variance taken at that location. The total number of observations in the initial experiment is $m_1 = n_1 \times r_0$. For the follow-up design e_2, assume that m_2 new observations are added, making the total number of observations $B = n_1 \times r_0 + m_2$. We denote the total number of distinct design points as $D = n_1 + n_2$. Hence, the new observations in the follow-up design can be divided as $m_2 = n_2 \times r_0 + \sum_{i=1}^{D} r_{si}$, where r_{si} indicates the additional replications taken at design point $\boldsymbol{x}_i$ and r_0 is the minimum number of replications assigned to every design point. As a result, the total number of replications at design location $\boldsymbol{x}_i$ after the follow-up design is $r_i = r_0 + r_{si}$.

The goal is then to select a follow-up design e_2 from a set of possible follow-up designs $\mathbf{E}_2$ to minimize the overall IMSPE. This design decision involves both determining the n_2 new design points and the additional r_{si} replicates to conduct

at all design points. As this involves looking at the design problem before the follow-up experiment is conducted, we take the prior expected value of the posterior variance. This is similar in spirit to the *preposterior* analysis in Raiffa & Schlaiffer (2000), Chapter 3.

Then, the design objective can be written as

$$\text{EIMSPE}(e_2^*) = \min_{e_2 \in \mathbf{E}_2} \int_{\boldsymbol{x}_0} \text{E}\left(\sigma^2_{|D_{n_1}, Y_{n_2}, \mathbf{S}^2_{n_2}, e_2}(\boldsymbol{x}_0)\right) d\boldsymbol{x}_0 \tag{6.10}$$

where e_2 and $Y_{n_2}, \mathbf{S}^2_{n_2}$ are included in the notation to indicate the additional data to be observed with the design e_2.

6.4.3 Simplification and decomposition of the IMSPE

To gain insights from this approach and criterion, we first look at the case when $\boldsymbol{\phi}$ and τ are known, as the predictive distribution in this case is tractable. As mentioned in Section 6.2.3.2, τ is the ratio of the error noise to the signal function, which can be dependent on the design locations. We let $\tau_i = \tau(\boldsymbol{x}_i)$ represent the ratio of the noise variability to the signal function at design location $\boldsymbol{x}_i$.

From Proposition 6.2.1, with τ known, the EIMSPE is given as

$$\begin{aligned} \text{EIMSPE}\,(e_2) &= \int_{\boldsymbol{x}_0} \text{E}\left[\sigma^2_{|Y_D, e_1, e_2}(\boldsymbol{x}_0)\right] dx_0 \\ &= \text{E}\left[\frac{2\gamma_z + Y_D^{\text{T}} R_D^{-1} Y_D + \mathbf{w}_0^{\text{T}} \mathbf{Q}_0^{-1} \mathbf{w}_0 - \boldsymbol{\lambda}_D^{\text{T}} \mathbf{M}_D^{-1} \boldsymbol{\lambda}_D}{2\alpha_z + n}\right] \times \\ &\quad \int_{\boldsymbol{x}_0} \left(1 - c_D^{\text{T}}(\boldsymbol{x}_0) R_D^{-1} c_D(\boldsymbol{x}_0) + \mathbf{h}_D \mathbf{M}_D^{-1} \mathbf{h}_D^{\text{T}}\right) d\boldsymbol{x}_0 \end{aligned} \tag{6.11}$$

where

$$\boldsymbol{\lambda}_D = F_D^{\text{T}} R_D^{-1} Y_D + \mathbf{Q}_0^{-1} \mathbf{w}_0, \;\; \mathbf{M}_D = F_D^{\text{T}} R_D^{-1} F_D + \mathbf{Q}_0^{-1}, \;\; \mathbf{h}_D = c_D^{\text{T}} R_D^{-1} F_D - F(\boldsymbol{x}_0),$$

$$Y_D^{\text{T}} = [Y_{n_1}^{\text{T}}, Y_{n_2}^{\text{T}}], \;\; c_D^{\text{T}} = [c_{n_1}^{\text{T}}, c_{n_2}^{\text{T}}], \;\; F_D^{\text{T}} = [F_{n_1}^{\text{T}}, F_{n_2}^{\text{T}}],$$

$$R_{n_1} = R_{zn_1} + R_{\xi n_1}, \;\; R_{n_2} = R_{zn_2} + R_{\xi n_2},$$

$$R_{\xi n_1} = \text{diag}\left[\tau_1/r_1, \tau_2/r_2, \ldots, \tau_{n_1}/r_{n_1}\right],$$

$$R_{\xi n_2} = \text{diag}\left[\tau_{n_1+1}/r_{n_1+1}, \tau_{n_1+2}/r_{n_1+2}, \ldots, \tau_{n_1+n_2}/r_{n_1+n_2}\right],$$

$$R_D = \begin{pmatrix} R_{n_1} & R_{n_1 n_2} \\ R_{n_1 n_2}^{\mathrm{T}} & R_{n_2} \end{pmatrix} = \begin{pmatrix} R_{zn_1} + R_{\xi n_1} & R_{n_1 n_2} \\ R_{n_1 n_2}^{\mathrm{T}} & R_{zn_2} + R_{\xi n_2} \end{pmatrix}.$$

Here, we note that the notation is written in terms of the number of distinct design points D. When the additional follow-up budget is only added at the existing n_1 design points, the sample means Y_{n_1} and R improves. To simplify the problem, here we assume that adding follow-up budget to the existing design points does not change the sample mean and the noise ratio τ_i, the correlation matrix R is affected only in the scale of the number of replications r_i. Therefore the EIMSPE criterion can be separately analyzed either in terms of the newly added design points $\boldsymbol{x}_{n_2}$ or the replication number r_i's. When only newly added distinct design points are considered, Y_{n_2} denotes the sample means of the simulation outputs at the n_2 new design points, R_{n_2} denotes the correlations between the new design points and $R_{n_1 n_2}$ denotes the correlations between the initial design points and the new design points.

Dividing Equation (6.11) into the following two terms:

$$s_0 = \mathrm{E}\left[\frac{2\gamma_z + Y_D^{\mathrm{T}} R_D^{-1} Y_D + \mathbf{w}_0^{\mathrm{T}} \mathbf{Q}_0^{-1} \mathbf{w}_0 - \boldsymbol{\lambda}_D^{\mathrm{T}} \mathbf{M}_D^{-1} \boldsymbol{\lambda}_D}{2\alpha_z + n}\right]$$

$$l_0 = \int \left(1 - c_D^{\mathrm{T}}(\boldsymbol{x}_0) R_D^{-1} c_D(\boldsymbol{x}_0) + \mathbf{h}_D \mathbf{M}_D^{-1} \mathbf{h}_D^{\mathrm{T}}\right) d\boldsymbol{x}_0$$

We see that s_0 represents the expected posterior mean of parameter σ_Z^2 given the vector Y_D of all observations (see Appendix D), and l_0 is a weighted function dependent on the locations of all design points and the prediction point $\boldsymbol{x}_0$. This weighted function represents the spatial influences of the design point locations on the prediction point.

As Equation (6.11) can be affected by increasing the number of replications at existing design points and also by exploring new design points, to illustrate insights on this experimental design problem in Stage 2, we consider two possible design options to improve the predictability of the metamodel for the two-point example:

Option 1 Adding replications to existing design points

Option 2 Adding new design points

6.4.3.1 A simplified Stage 2 design for the two point example

To gain useful insights into the design issue, we adopt the two point example in Section 6.2.3.1 to illustrate the influence of the different design options. To further simplify the analysis, we focus on the influence of the options on l_0. This is similar to using another common design criterion, the standardized IMSPE criterion

$$\text{IMSPE}_S = \int_{\boldsymbol{x}_0} \frac{\sigma^2_{|Y_{n_1}}(\boldsymbol{x}_0)}{\sigma^2_Z} d\boldsymbol{x}_0 \tag{6.12}$$

proposed by Sacks *et al.* (1989) and also adopted in Santner *et al.* (2003) (Chapter 6). With this, the EIMSPE in Equation (6.11) reduces to the weight function l_0.

Adopting the design setting in Section 6.2.3.1 where observations y_1 and y_2 are obtained at the initial design points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, which are a distance of d_0 apart. Suppose the τ values and number of replications at each design point are τ_1, τ_2 and r_0. Here we also assume that the prediction point $\boldsymbol{x}_0$ is located in the middle of the two design points, and adopt the linear correlation function, $\rho_0 = 1 - \phi_Z|d_0|$ where $d_0 = \boldsymbol{x}_1 - \boldsymbol{x}_2$, and the constant underlying mean function.

Suppose after the initial kriging model is fit based on this data, additional computing budget r_0 is available. Here we consider two different design options for this additional budget: option 1 as adding additional replications r_{s1} and r_{s2} to existing design points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ with the constraint that $r_{s1} + r_{s2} = r_0$ and option 2 as adding a new design point $\boldsymbol{x}_3$ in between points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

In order to compare the influences of the two design options on the weight function l_0, we use the reduction on l_0, $\Delta = l_0 - l_0'$, as the evaluation measure, where l_0' represents the value of the weight function with the follow-up design. l_0' can have different values with different design options used.

For design option 1 of adding additional replications, the new correlation matrix can be given as

$$R_D = R' = \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{\tau_1}{r_0+r_{s1}} & 0 \\ 0 & \frac{\tau_2}{r_0+r_{s2}} \end{pmatrix}$$

Then the reduction $\Delta_1 = l_0 - l_0'$ will equal to

$$\begin{aligned}\Delta_1 &= \left[\frac{(r_0+r_{s2})(1-\rho_0)(2(r_0+r_{s1})(1-\rho_0)3\tau_1)}{(r_0+r_{s2})(2(r_0+r_{s1})(1-\rho_0)+\tau_1)+(r_0+r_{s1})\tau_2}\right. \\ &+ \frac{3\tau_2((r_0+r_{s1})(1-\rho_0)+\tau_1)}{(r_0+r_{s2})(2(r_0+r_{s1})(1-\rho_0)+\tau_1)+(r_0+r_{s1})\tau_2} \\ &+ \left.\frac{2r_0^2(1-\rho_0)^2+3\tau_1\tau_2+3r_0(1-\rho_0)(\tau_1+\tau_2)}{r_0(2r_0(1-\rho_0)+\tau_1+\tau_2)}\right]\frac{1-\rho_0}{2} \end{aligned} \tag{6.13}$$

The first order derivative of Equation (6.13) with respect to r_{s1} and r_{s2} are positive functions,suggesting that adding additional replications can reduce the value of l_0 and the reduction is a monotonic increasing function in the number of the additional replications r_{s1} and r_{s2} added.

For design option 2 of adding new design points, assume that the new point added is $\boldsymbol{x}_3$, and let d_t denote the distance between point $\boldsymbol{x}_1$ and $\boldsymbol{x}_3$.

For computational simplicity, we reparameterize the correlation between $Z(\boldsymbol{x}_1)$ and $Z(\boldsymbol{x}_3)$ as $\rho_t = 1 - \phi|d_t|$. The correlation matrix can then be written as

$$R_D = R' = \begin{pmatrix} 1 & \rho_t & \rho_0 \\ \rho_t & 1 & 1+\rho_0-\rho_t \\ \rho_0 & 1+\rho_0-\rho_t & 1 \end{pmatrix} + \begin{pmatrix} \frac{\tau_1}{r_0} & 0 & 0 \\ 0 & \frac{\tau_3}{r_0} & 0 \\ 0 & 0 & \frac{\tau_2}{r_0} \end{pmatrix}.$$

The reduction $\Delta_2(\rho_t) = l_0(\rho_t) - l_0'(\rho_t)$ will then be

$$\begin{aligned}\Delta_2(\rho_t) &= ((2r_0(1-\rho_t)+\tau_1)^2((2r_0(1-\rho_0)+\tau_2)^3-\tau_2^3)) \\ &/ \quad (6r_0^2(2r_0(1-\rho_0)+\tau_1+\tau_2) \\ &\quad ((2r_0(1-\rho_t)+\tau_1)(2r_0(\rho_t-\rho_0)+\tau_2)+\tau_3(2r_0(1-\rho_0)+\tau_1+\tau_2)))\end{aligned}$$

As seen, $\Delta_2(\rho_t)$ is a positive function, which suggests that adding new design points into the existing design can reduce the value of l_0. The reduction is a function of ρ_t, where $\rho_t \in [\rho_0, 1]$. Hence, it is possible to optimize $\Delta_2(\rho_t)$ in terms of ρ_t.

Since the τ_1, τ_2, τ_3 are unknown, we assume certain structures for τ_1, τ_2, τ_3 to reflect some general homoscedastic and heteroscedastic cases.

Assuming $\tau_1 = \tau_2 = \tau_3 = \tau$ for the homoscedastic case, the correlation matrix becomes

$$R' = \begin{pmatrix} 1 & \rho_t & \rho_0 \\ \rho_t & 1 & 1+\rho_0-\rho_t \\ \rho_0 & 1+\rho_0-\rho_t & 1 \end{pmatrix} + \begin{pmatrix} \frac{\tau}{r_0} & 0 & 0 \\ 0 & \frac{\tau}{r_0} & 0 \\ 0 & 0 & \frac{\tau}{r_0} \end{pmatrix}.$$

Maximizing the reduction function $\Delta_2(\rho_t)$ gives $\rho_t^* = \frac{1+\rho_0}{2}$ and $\Delta_2^*(\rho_t) = \frac{(1-\rho_0)(r_0^2(1-\rho_0)^2+3r_0(1-\rho_0)\tau+3\tau^2)}{6r_0(r_0(1-\rho_0)+3\tau)}$. Comparing $\Delta_2^*(\rho_t)$ and Δ_1^* (obtained by maximizing Δ_1 in Equation (6.13) with respect to r_{s1} and r_{s2}), we see that $\Delta_2^*(\rho_t) \geq \Delta_1^*$ when

$$\begin{aligned}\tau \;&\leq\; (-2^{2/3}21^{1/3}(12(27r_0^3(1-\rho_0)^3+\sqrt{57r_0^6(1-\rho_0)^6})^{1/3})^{2/3}\\ &+\; 18r_0(12(27r_0^3(1-\rho_0)^3+\sqrt{57r_0^6(1-\rho_0)^6})^{1/3})^{1/3}(1-\rho_0)\\ &-\; 4\times 2^{1/3}21^{2/3}r_0^2(1-\rho_0)^2)/(12(27r_0^3(1-\rho_0)^3+\sqrt{57r_0^6(1-\rho_0)^6})^{1/3})\end{aligned}$$

suggesting that adding new points is better than adding additional replications for this homoscedastic case when the noise level is lower than a certain threshold.

For the heteroscedastic case, we assume that the noise function is a linear function of the location of the design point, with $\tau_1 = 0, \tau_2 = \tau, \tau_3 = (1-\rho_t)/(1-\rho_0) * \tau$. The correlation matrix becomes

$$R' = \begin{pmatrix} 1 & \rho_t & \rho_0 \\ \rho_t & 1 & 1+\rho_0-\rho_t \\ \rho_0 & 1+\rho_0-\rho_t & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{(1-\rho_t)}{(1-\rho_0)} * \frac{\tau}{r_0} & 0 \\ 0 & 0 & \frac{\tau}{r_0} \end{pmatrix}.$$

Maximizing the reduction function $\Delta_2^*(\rho_t)$, we can obtain a closed form solution for ρ_t^* and $\Delta_2^*(\rho_t)$. Setting $r_0 = 2, d_0 = 1, \phi_z = 0.5$, the correlation is $\rho_0 = 1-\phi_z|d_0| = 0.5$. Figure 6.4 plots how the change of noise level τ will affect the optimal location of the new design point. As seen in Figure 6.4, the value of d_t^* increases as the noise level τ increases, suggesting that the optimal location of the new design point will be closer to the design point $\boldsymbol{x}_2$ with higher variability.

Figure 6.5 plots $\Upsilon = \Delta_2^* - \Delta_1^*$ with τ for $r_0 = 2$ and different values of ϕ_z. From this figure, we see that Υ is a monotonic decreasing function of τ. When $\phi_z = 0.5$, $\Upsilon = 0$ at $\tau^* = 6.31$, suggesting that when $\tau < 6.31$, adding new design points is better than adding new replications into the existing design. When $\tau > 6.31$, adding additional replications will be the better option (with the optimal allocation assigning additional replications to $\boldsymbol{x}_2$). It can further be shown that τ^* is a monotonic decreasing function of ρ_0, which suggests that adding new design points to the existing design will have better reductions as the spatial correlation becomes weaker. This is somewhat intuitive as when the

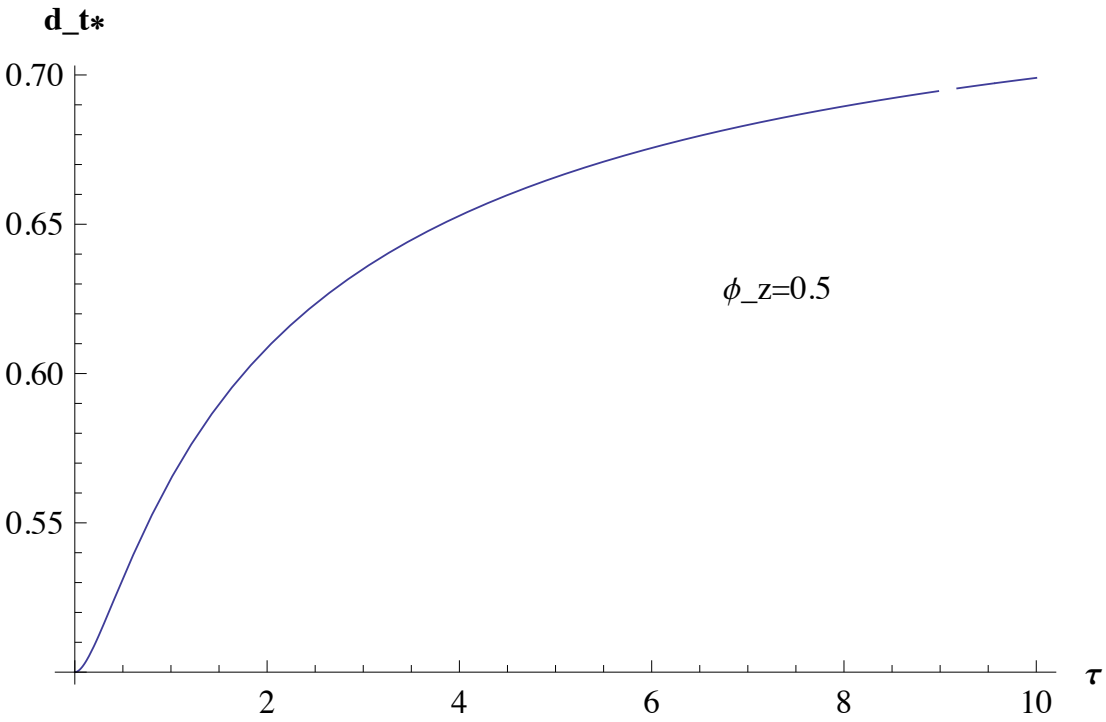


Figure 6.4: The influence of the random noise level on the optimal location of the new design point for the two point case with $\phi_z = 0.5$.

spatial correlations are high, inference on points in between design points can be made with existing design points, hence adding additional replications (Option 1) at the existing design points to improve its estimate provides better reduction in the EIMPSE$_s$ when τ becomes slightly large. However, when the design points have weak or no spatial correlations, which suggest an increasing value for ϕ_z, an intermediate design point, closer to the prediction point, is desired to provide more information at the prediction point.

6.4.3.2 A numerical study on the EIMSPE for different design options

Here, to study the effects of the design options on the EIMSPE criterion given in Equation (6.11), consider a simple linear function $y = x + \varepsilon(x)$, where $x \in [0, 1]$. The initial design is a uniform design located in $[0, 1]$. We vary the number of initial design points, the level of variability of $\varepsilon(x)$ and the functional form of $\varepsilon(x)$. The details of various test scenarios are given in Table 6.5. For all the scenarios,

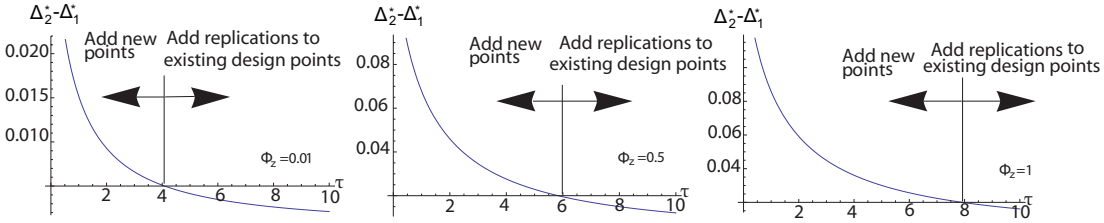


Figure 6.5: Υ with $\phi_z = 0.01, 0.5, 1$.

Table 6.5: Setting of Different Test scenarios.

No. of scenario	n_1	variance level	variance type
1	3	1	constant
2	7	1	constant
3	3	100	constant
4	7	100	constant
5	3	$\sigma_\varepsilon^2 = x$	variable
6	7	$\sigma_\varepsilon^2 = x$	variable
7	3	$\sigma_\varepsilon^2 = 2x$	variable
8	7	$\sigma_\varepsilon^2 = 2x$	variable

we set the number of replications taken at each design point at $r_0 = 1000$. We consider the design allocation of an additional 1000 runs on either design Option 1 (where the additional runs are distributed to the existing points) or design Option 2 (where a new design point is selected with 1000 replications run at that point). 100 possible new design points for the follow-up design are uniformly placed from 0 to 1. These possible new design points are placed such that they do not overlap with the existing design points. Figure 6.6 plots the EIMSPE when the two different design options are applied to the eight test scenarios in Table 6.5. The dotted line shows the EIMSPE for different new points x over the design space (to select optimal new design point for Option 2). The solid line indicates the optimal EIMSPE value for the optimal allocation of replications to existing design points (Option 1), drawn here as a line for comparison with Option 2. Some interesting characteristics of the plots are summarized below:

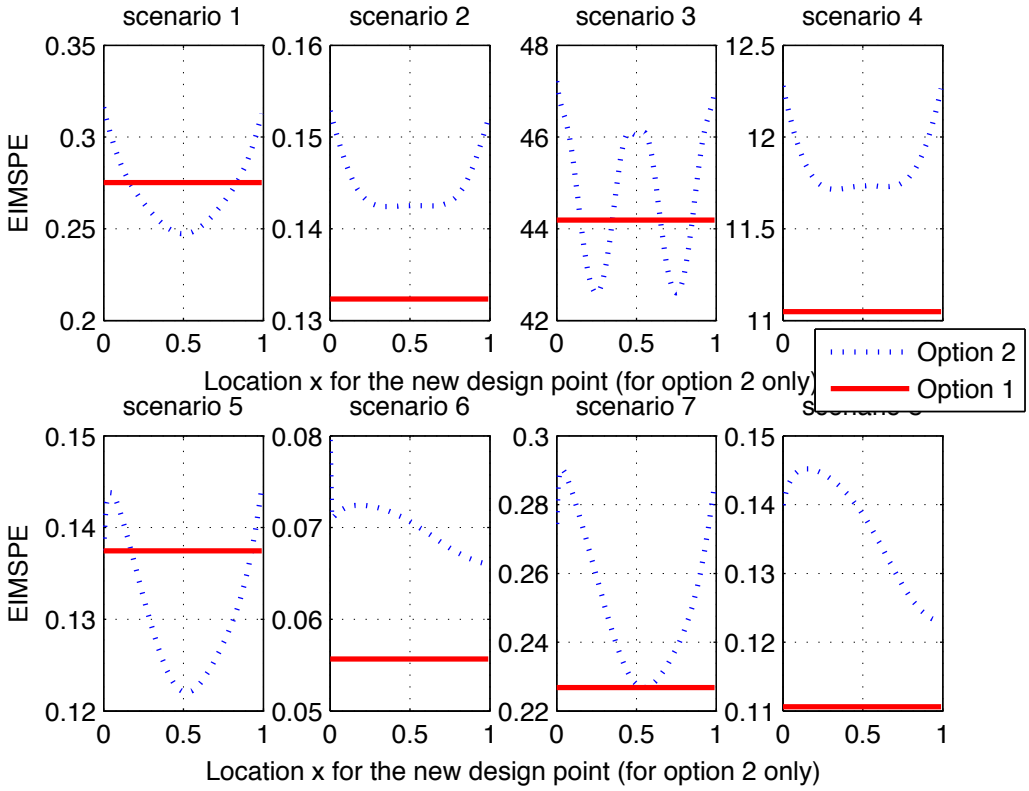


Figure 6.6: Results for Option 1 (solid line) and Option 2 (dotted line) for the eight test scenarios.

1 When noise is small and the initial design is sparse, there is tendency to place new design points close to the center of the design space (see scenario 1). In this low variance case, spatial correlations are strong, giving centrally located design points the most influence. This occurs also in the low variance heteroscedastic case (see scenario 5).

2 When noise is moderately small but heteroscedastic and the initial design points are sparse (see scenario 7), new design points are still focused in the center of the design region. Placing additional replications at the existing center point is as effective too.

3 When the noise is large and the initial design is sparse (as in scenario 3), the spatial correlation is weaken by the poor estimation and high variability.

The design criterion will tend to place new design point in between initial design points, suggesting that the follow-up design will approach a space filling design.

4 When the initial design points sufficiently fills the design space (as scenarios 2, 4, 6, 8), adding more replications to existing design points (Option 1) to reduce the variance at these points are preferred.

5 Considering only Option 2 in these scenarios, when the variance is constant (scenarios 2 and 4), new design points in the central region of the design space is favored. In the heteroscedastic case (scenarios 6 and 8), new design points are placed in the region of higher variance.

6 Consider only Option 1 for all, when the variance is homogenous (scenarios 1-4), additional replications are focused on existing design points that are centrally located in the design space. In the case of heteroscedastic variance, additional replications are allocated at the points with the highest variance.

Overall, some insights and suggestions can be drawn from these results:

1 When the design space is sufficiently filled, in all cases, additional replications are added to existing design points to reduce variability. These additional replications are focused in centrally located regions and in regions with higher variability.

2 When the initial design space is sparse, preference is for allocating new design points in regions where spatial influence is strong. However, the effect of spatial influence is dependent on the noise level too, where the spatial relations are weakened by higher noise. In situations where noise is high, denser designs are favored.

Based on the results and summary of the eight test scenarios, there are several different factors that can affect the EIMSPE. Moreover, the definition of filled/sparse is dependent on the problem situation, the noise level and also the correlation functions assumed. Hence, it is difficult to balance all these factors at the same time in the experimental design, especially with limited computing

budget. In order to solve the problem of balancing between adding new follow-up design points and adding additional replications to the current design points with limited computing budget, we propose two possible design approaches based on the EIMSPE given in Equation (6.10) in this chapter.

6.4.4 Improved two-stage design approaches

Here we assume that the total computing budget is B runs. We further assume that the minimum number of replications for each design point is r_0. Suppose the initial design has n_1 unique design points, then the follow-up design problem is to determine the allocation of the remaining $m_2 = B - n_1 \times r_0$ runs. As mentioned previously, the general two-stage framework can be modified in order to allocate the remaining computing budget more efficiently. We propose two different approaches in this section that accounts for different design scenarios.

6.4.4.1 One-Point-at-A-Time (OPAT) sequential design approach

In this approach, since a minimum of r_0 replications is required at each design point, we allocate r_0 runs at each step. This is continued until the computing budget is exhausted. With this allocation procedure, at each step, the experimenter has two decision options: (1) to optimally distribute the r_0 runs to the existing design points as additional replications or (2) to assign all r_0 runs to a new design point. For option (1), the additional replications should be optimally distributed among all the current design points to minimize the EIMSPE in Equation (6.10). For option (2), the new design point $\boldsymbol{x}_{n_2}$ should be optimally chosen to minimize the EIMSPE.

Here we consider the M/M/1 queue example with a total computing budget of $B = 1000$ and the minimum number of replication for each point $r_0 = 100$. 3 initial design points are first located at $x = 0.50, 0.675, 0.85$. For this OPAT approach, in the follow-up design, there are $(B - r_0 \times n_1)/r_0 = 7$ steps. We consider applying both the BKS simplification and the fully Bayesian MCMC.

Table 6.6 and Table 6.7 outlines the steps 1 to 7 taken by the algorithm. With the BKS model (Table 6.6), the algorithm quits with final design points located at $x = 0.50, 0.565, 0.652, 0.675, 0.756, 0.818, 0.85$ and the final replications

distributed to each point are $100, 131, 201, 127, 179, 120, 142$, respectively. The final design has 7 design points (3 initial and 4 new), and the EIMSPE of 0.0854. With the BKMCMC model (Table 6.7), the algorithm quits with final design points located at $x = 0.50, 0.599, 0.675, 0.702, 0.804, 0.85$ and final replications distributed to each point are $102, 162, 220, 157, 231, 128$, respectively. The final design has 6 design points (3 initial and 3 new), and the EIMSPE is 0.1277. As the BKS model assumes more initial information (parameters known), the final design from this approach has a smaller EIMSPE.

Table 6.6: Step by step allocation with the BKS model († indicates initial design point, *indicates new design point in this step)

Design point (x)	$0.5^{\dagger}$	0.565	0.652	$0.675^{\dagger}$	0.756	0.818	$0.85^{\dagger}$
Initial design (reps)	100			100			100
Step 1*			100				
Step 2*					100		
Step 3*		100					
Step 4	0	11	45	9	27		8
Step 5*						100	
Step 6	0	10	28	9	26	6	21
Step 7	0	10	28	9	26	14	13
Total reps	100	131	201	127	179	120	142

The location of the design points when using BKS and BKMCMC are similar, although the BKMCMC design has 1 less design point and more replications allocated at each design point. As the BKMCMC model accounts for more parameter uncertainties, the variability is higher (higher EIMSPE), the algorithm places more replications at each point after the design points have been spaced out in the design region. Overall, the OPAT can balance the needs of adding new design points and additional replications to the existing design points, which is especially useful for the heteroscedastic case.

Table 6.7: Step by step allocation with the BKMCMC model († indicates initial design point, *indicates new design point in this step)

Design point (x)	$0.5^{\dagger}$	0.599	$0.675^{\dagger}$	0.702	0.804	$0.85^{\dagger}$
Initial design (reps)	100		100			100
Step 1*				100		
Step 2*		100				
Step 3*					100	
Step 4	0	16	30	14	34	6
Step 5	0	16	30	15	32	7
Step 6	1	15	30	14	33	7
Step 7	1	15	30	14	32	8
Total reps	102	162	220	157	231	128

6.4.4.2 Simple two-stage design approach

In a strictly two-stage design approach, all the remaining computing budget $m_2 = B - n_1 \times r_0$ are to be distributed in the second stage at one time. In this approach, the experimenter still needs to balance the computing budget between adding new design points and additional replications to the existing design points. Let e_2 be the follow-up design which includes $\boldsymbol{X}_{n_2}$, the location matrix of the follow-up design points and r_{si}, the number of additional replications for each design point i, $i = 1, \ldots, D$, where $D = n_1 + n_2$ and n_2 is the number of follow-up design points in $\boldsymbol{X}_{n_2}$. Let $\boldsymbol{\Xi}$ be the finite design space of all possible combination of inputs for each of the n_2 possible runs. The design problem can be formulated as

$$\min_{r_{si} \in \mathbb{Z}, n_2 \in N, \boldsymbol{X}_{n_2} \in \boldsymbol{\Xi}} [\text{EIMSPE}(\boldsymbol{X}_{n_2}, r_{si})] \tag{6.14}$$

subject to

$$N \in \mathbb{Z}, N \leq m_2/r_0, \sum_{i=1}^{D} r_i = B,$$

where $r_i = r_0 + r_{si}$.

To solve this problem, we first fix n_2, then optimally distribute all the remaining computing budget. We repeat this for all $n_2 = 0, 1, \ldots, \lfloor m_2/r_0 \rfloor$.

The basic procedures to this simple two-stage design approach can be summarized as:

1 In the first stage, select a space-filling design as the initial design with n_1 design points.

2 In the second stage, to solve Equation (6.14), enumerate all the possible cases with different number of newly added follow-up design points $n_2 = 0, 1, \ldots, \lfloor m_2/r_0 \rfloor$. For each case, the locations of the new design points $\boldsymbol{X}_{n_2}$ should be optimally chosen to minimize the EIMSPE in Equation (6.11). After that, all the remaining computing budget $B-(n_1+n_2)\times r_0$ is optimally allocated among all the $n_1 + n_2$ design points.

Here we use the M/M/1 queue example again to illustrate this approach. Consider an initial design with 5 design points located at $x = 0.50, 0.5875, 0.675, 0.7625, 0.85$.

The total additional computing budget $B - n_1 * r_0 = 500$ and the minimum number of replication for each point is $r_0 = 100$. Applying the solution procedure, Table 6.8 shows the best design for $n_2 = 0, \ldots, 5$ when using the BKS and BKMCMC models. The design results provided by the BKMCMC model are different from the results provided with the BKS model. Furthermore, as the BKMCMC model accounts for all the parameter estimation uncertainties, the variability of its predictive distribution is higher than that of the BKS model. As a result, the EIMSPE provided with the BKMCMC model is higher. The optimal design is selected as the one with the lowest EIMSPE. For the BKS model, the optimal design is the best design obtained when two new design points are added ($n_2 = 2$). The optimal design obtained when BKMCMC is used is the design with no new points added ($n_2 = 0$). The number of additional replications allocated in this case is increasing towards the high traffic regions, with three times more runs at $x = 0.85$ than at $x = 0.5$. This result is similar to the results in Section 6.4.4.1 and is intuitively related to the conclusions in Section 6.4.3, which suggests that the experimenter should focus on reducing the variability of the observed data by adding more replications to the existing design point when the variability is high and the design space is sufficiently filled.

Table 6.8: Design results from the simple two-stage design approach. (* *indicates optimal design*)

n_2	BKS EIMSPE	BKMCMC EIMSPE
0	0.170	0.206*
1	0.382	0.638
2	0.083*	0.375
3	0.085	0.432
4	0.154	0.491
5	0.147	0.325

This simple two-stage design approach is suitable for batch type simulation experiments, where computer codes are executed in batch modes or in parallel computing environments.

6.5 Comments and Conclusions

Kriging is an increasingly popular metamodeling tool in simulation due to its flexibility in global fitting and prediction. In the fitting of this metamodel, the parameters are likely to be estimated from the simulation data and traditional plug-in estimators for prediction typically ignore these parameter estimation uncertainties. These can be substantial in stochastic simulations and can lead to overconfident results and misguided decisions. In this chapter, we propose a Bayesian kriging metamodeling approach for stochastic simulation that can account for all parameter uncertainties as well as the inherent uncertainty of the simulation model itself, providing a more accurate approach for prediction and analyses. We derive the predictive distribution under certain distribution assumptions and proposed a MCMC approach to address the general case. The results in Section 6.2 and Section 6.3 suggest that this proposed Bayesian approach can better account for the noise and heterogeneity in stochastic simulations. In addition, we further consider the important problem of planning the experimental

design. We propose a two stage design approach that systematically balances the allocation of computing resources to new design points and replication numbers in order to reduce the uncertainties and improve the accuracy of the predictions. We illustrate the design approach with several examples and provide some general suggestions on design plans under various scenarios. The current work can be extended in several directions. First, as observed in Section 6.3, selection of the prior mean functional form is important, especially in the presence of large noise. One possibility is to include model selection techniques within the modeling framework to guard against model misspecification. Another possibility is to include this model uncertainty into the framework. Second, the proposed design criterion focuses on improving the overall predictive ability of the metamodel. As metamodels have also been widely used for simulation optimization, our proposed metamodeling approach can be integrated into optimization techniques to develop sequential approaches for global optimization. For example, the predictor and its distribution form as described in Equation (6.5) can be substituted into optimization search point criterions like the expected improvement (EI) function (Jones *et al.* (1998)) or augmented forms of the EI function (Huang *et al.* (2006)) and applied in optimization algorithms like Efficient Global Optimization (EGO) for global optimization.

Chapter 7

CONCLUSION

This study contributes to the design and analysis of computer experiment for stochastic systems. More specifically, we investigate both the modeling and experimental design in the stochastic simulation context. In this chapter, we are going to conclude the study by summarizing the findings and limitation of this research, and also the possible future research topics.

7.1 Main findings

In this thesis, we first investigate the kriging model's application in stochastic simulation, especially in heteroscedastic situations in Chapter 3. We discuss the kriging model's behavior given the stochastic simulation outputs, and propose a modified nugget-effect model by relaxing the stationarity assumptions of the nugget-effect model and modeling the random noise as an independent random process. We then look into the behavior of the likelihood function in the parameter estimation for the traditional OK model given the stochastic inputs, and note that the nugget-effect model and the modified nugget-effect model can reduce the erratic behavior in the parameter estimation by penalizing the likelihood function affected by noisy input. This modified nugget-effect model is also compared with the stationary kriging model in terms of the mean squared error. We propose three methods to easily obtain the prediction variance of our new predictor; namely, the kriging modeling, non-parametric bootstrapping method and the interpolation method. We illustrate our model and approach with several

numerical examples. Our results indicate that the proposed modified nugget-effect model is a promising method for applications in simulation systems with heterogeneous variances. In the following Chapter 4, we investigate the impact of parameter estimation uncertainty on three different kriging model forms in stochastic simulations. We analyzed theoretically a simple two-point problem and conducted three numerical studies. Based on the results of these studies, we find that the sensitivity parameter ϕ_Z estimated by kriging model is affected by the random noise in the stochastic system, and the additional prediction error caused by parameter estimation uncertainty increases as the variance of the random noise $\varepsilon(x)$ increases. The proportion of the additional prediction error caused by parameter estimation uncertainty in overall prediction error increases when the variance of the random noise $\varepsilon(x)$ increases. In the case when the variability of the noise is low and sufficient sample data is available, this additional error becomes negligible. Among the three kriging model forms studied in this chapter, the modified nugget effect model seems to have the best performance in both the overall prediction error and additional prediction error caused by parameter estimation uncertainty. This phenomenon is partially explained in the two-point problem adopted in this chapter.

We further apply the modified nugget effect kriging model to the stochastic simulation optimization problem in the Chapter 5. With the modified nugget effect kriging model, we adopt the efficient global optimization framework and expected improvement function to allocate the available computing budget for the experimental design. Hence the traditional design methods can not balance the different design options in the scenarios with heterogeneous variance, we propose the two-stage sequential design framework to solve the problem. The proposed design framework tends to reduce the random variability in the first stage to certain level and then search for the new design point in the second stage. We test the proposed design framework with a simple one dimensional example then followed by a more realistic shipping liner service planning simulation example. The results suggest that the proposed two-stage sequential design framework is a more reasonable design method to be used for the stochastic simulation than the traditional method.

After the development of the kriging model with modified nugget effect and related experimental design methods, we realize that the parameter estimation uncertainty problem mentioned in Chapter 4 still needs appropriate solution. In Chapter 6, we propose a Bayesian kriging metamodeling approach for stochastic simulation that can account for all parameter uncertainties as well as the inherent uncertainty of the simulation model itself, providing a more accurate approach for prediction and analysis. We derive the predictive distribution under certain distribution assumptions and proposed a MCMC approach to address the general case. The results in Section 6.3 suggest that this proposed Bayesian approach can better account for the noise and heterogeneity in stochastic simulations. In addition, we further consider the important problem of planning the experimental design. We propose a two stage design approach that systematically balances the allocation of computing resources to new design points and replication numbers in order to reduce the uncertainties and improve the accuracy of the predictions. We illustrate the design approach with several examples and provide some general suggestions on design plans under various scenarios.

7.2 Future research

The current work can be extended in several directions.

1. For the optimization method in Chapter 5, developing adaptive schemes that dynamically distributes the budget for each iteration, studying in detail the convergence results of the algorithm and comparisons with other simulation optimization approaches.

2. As observed in Section 6.3.1, selection of the prior mean functional form is important, especially in the presence of large noise. One possibility is to include model selection techniques within the modeling framework to guard against model misspecification. Another possibility is to include this model uncertainty into the framework.

3. The computational efficiency of the current proposed Bayesian Kriging still needs to be improved, especially for the sampling methods for the unknown parameters in the MCMC algorithm

4. For the matrix computation simplification issue of the kriging metamodel, both the accuracy and efficiency are to be carefully considered. Due to the large number of matrix inverse calculations needed in the application of kriging model, the model is not able to handle the large size inputs. Hence alternative methods can be applied to reduce the complexity of the computation.

References

A. Sobester, S.L. & Keane, A. (2002). The use of surrogate modeling algorithms to exploit disparities in function computation time within simulation-based optimization. In *In Proceedings of the Third ISSMO/AIAA Internet Conference on Approximations in Optimization*, 427–430, ISSMO/AIAA. 28

Agresti, A. (2002). *Categorical Data Analysis*. Wiley, New York, 2nd edn. 180

Angün, E., Hertog, J.P. & Gürkan, G. (2002). Response surface methodology revised. In *Proceedings of the 2002 Winter Simulation Conference*, 377–383, Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. 6, 84

Ankenman, B., Nelson, B.L. & Staum, J. (2010). Stochastic kriging for simulation metamodeling. *Operations Research*, 371–382. 21, 30, 34, 41, 49, 94, 115, 117, 119, 120, 122, 128, 130, 133

Anker, L. & Jurs, P. (1992). Prediction of nuclear-magnetic-resonance chemical-shifts by artificial neural netwoks. *Analytical Chemistry*, **64**, 1157–1164. 17

Asmussen, S. & Glynn, P.W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer, New York, 1st edn. 5

Bakin, S., Hegland, M. & Osborne, M. (1992). Parallel mars algorithm based on b-splines. *Computational Statistics*, **15**, 463–484. 15

Barton, R.R. (1992). Metamodels for simulation input-output relations. In J.J. Swain, D. Goldsman, R.C. Crain & J.R. Wilson, eds., *Proceedings of the 1992 Winter Simulation Conference*, 289–299, Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. 19

Barton, R.R. (1998). Simulation metamodels. In D.J. Medeiros, E.F. Watson, J.S. Carson & M.S. Manivannan, eds., *Proceedings of the 1992 Winter Simulation Conference*, 167–174, Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. 19

Beers, W.V. & Kleijnen, J. (2003). Kriging for interpolation in random simulation. *Journal of the Operational Research Society*, **54**, 255–262. 18, 20, 21, 33

Bekki, J.M., Fowler, J.W., Mackulak, G.T. & Andkulahci, M. (2009). Simulation-based cycle-time quantile estimation in manufacturing settings employing non-fifo dispatching policies. *Journal of Simulation*, **3**, 69–83. 122

Box, G. & Draper, R. (1987). *Empirical Model-Building with Response Surface*. Wiley, New York. 25

Celis, M., Dennis, J. & Tapia, R. (1984). A trust region strategy for nonlinear equality constrained optimization. In R.B.S. Paul T. Boggs Richard H. Byrd, ed., *In Proceedings of the SIAM Conference on Numerical Optimization*, 71–82, SIAM, Boulder, Colorado. 26

Chen, C., Lin, J., Yücesan, E. & Chick, S.E. (2000). Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dynamic Systems: Theory and Applications*, **10**, 251–270. 92, 94, 177

Chen, X., Ankenman, B. & Nelson, B.L. (2012). The effects of common random numbers on stochastic kriging metamodels. *ACM Transactions on Modeling and Computer Simulation*, **22**. 21

Cheng, R. & Kleijnen, J. (1999a). Improved design of queueing simulation experiments with highly heteroscedastic responses. *Operations Research*, **47**, 762–777. 13, 122

Cheng, R.C.H. & Kleijnen, J.P.C. (1999b). Improved design of queueing simulation experiments with highly heteroscedastic responses. *Operations Research*, **47**, 762–777. 59

Chin, D. & Melone, F. (1999). Using airspace simulation to assess environmental improvements from free flight and cns/atm enhancements. In P.A. Farrington, H. Nembhard, D. Sturrock & G.W. Evans, eds., *In Proceedings of the 1999 Winter Simulation Conference*, 1295–1301, Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. 2

Cowles, M.K. & Carlin, B.P. (1996). Markov chain monte carlo convergence disgnostics: A comparative review. *Journal of the American Statistical Association*, **91**, 883–904. 127

Cressie, N. (1993). *Statistics for spatial data*. Wiley, New York, revised edn. 6, 7, 18, 19, 30, 35, 38, 39, 42, 47, 67, 68, 69, 72, 86, 119, 170

Currin, C., Mitchell, T., Morris, M. & Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Associaiton*, **86**, 953–963. 1, 18, 120

den Hertog, Kleijnen, J.P.C. & Siem, A.Y.D. (2005). The correct kriging variance estimated by bootstrapping. *Journal of the Operational Research Society*, **165(3)**, 826–834. 50, 51, 68, 69, 117, 122

Dyn, N. & Yad-Shalom, I. (1991). Optimal distribution of knots for tensor product spline approximations. *Quarterly for Applied Mathematics*, **49**, 19–27. 15

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7**, 1–26. 50

Fang, K., Li, R. & Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC, Boca Raton. 24

Fang, K.T. (1980). The uniform desgin: application of number-theoretic methods in experimental design. *Acta Mathematicae Applicatae Sinica*, **3**, 363–372. 24

Fang, K.T., Lin, D.K., Winker, P. & Zhang, Y. (2000). Uniform design: Theory and application. *Technometrics*, **42**, 237–248. 24

Friedman, J. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, **19**, 1–141. 14

Funahashi, K.I. (1989). On the approximate realization of continuous mappings by neural networks. *Nerual Networks*, **2**, 183–192. 17

Gano, S., Renaud, J., Martin, J. & Simpson, T. (2006). Update strategies for kriging models used in variable fidelity optimization. *Structural and Multidisciplinary Optimization*, **32**, 287–298. 27

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2003). *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, 2nd edn. 124, 127

Gramacy, R.B. & Lee, K.H. (2009). Adaptive design and analysis of supercomputer experiments. *Technometrics*, **51**, 130–145. 32

Greenwood, A., Vanguri, S., Eksioglu, B., Jain, P., Hill, T., Miller, J. & Walden, C. (2005). Simulation optimization decision support system for ship panel shop operations. In *Proceedings of the 2005 Winter Simulation Conference*, 2078–2086, Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. 2

Gupta, A., Ding, Y., Xu, L. & Reinikainen, T. (2006a). Optimal parameter selection for electronic packaging using sequential computer simulations. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, **128**, 705–715. 30

Gupta, A., Yu, D., Xu, L. & Reinikainen, T. (2006b). Optimal parameter selection for electronic packaging using sequential computer simulations. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, **128**, 705–715. 17, 32

Hardy, R. (1971). Multiquadratic equations of topography and other irregular surface. *Journal of Geophysical Research*, **76**, 1905–1915. 15

Harville, D. (1985). Decomposition of prediction error. *Journal of American Statistical Association*, **80**, 132–138. 68, 75

Harville, D. & Jeske, D.R. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of American Statistical Association*, **87**, 724–731. 68

Hillier, F.S. & Lieberman, G.J. (2001). *Introduction to Operation Research*. McGraw-Hill, 7th edn. 57

Hornik, K., Stinchcombe, M. & White, H. (1989). Multilayer feedforward networks are universal approximators. *Nerual Networks*, **2**, 359–366. 17

Hsu, K., Gupta, H. & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resource Research*, **31**, 2517–2530. 17

Huang, D., Allen, T.T., Notz, W.I. & Zeng, N. (2006). Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization*, **34(3)**, 441–466. 6, 19, 28, 30, 61, 85, 86, 94, 151

Hussain, M., Barton, R. & Joshi, S. (2002a). Metamodeling: radial basis function, versus polynomials. *European Journal of Operational Research*, **138**, 142–154. 16

Hussain, M.F., Barton, R.R. & Joshi, S.B. (2002b). Metamodeling: radial basis functions, versus polynomials. *European Journal of Operational Research*, **138**, 142–154. 76, 79

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, London. 123

Jin, R., Chen, W. & Simpson, T. (2000). Comparative studies of metamodeling techniques under multiple modeling criteria. In *In Proceedings of the 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 4801, AIAA, Long Beach, CA. 15

Johnson, M.E., Moore, L.M. & Ylvisaker, D. (1990). Minimax and maxmin distance design. *Journal of Statistical Planning Inference*, **26**, 131–148. 25

Jones, D. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, **21**, 345–383. 85, 86

Jones, D., Schonlau, M. & Welch, W. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, **13**, 455–492. 5, 27, 84, 86, 96, 98, 100, 151

Kackar, R. & Harville, D. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of American Statistic Association*, **79**, 853–862. 68, 70

Kansa, E. (1990). Multiquadrics-a scattered data approximation scheme with applications to computation fluiddynamics-i: surface approximations and partial derivatives estimates. *Computers and Mathematics with Applications*, **19**, 127–145. 16

Kennedy, M. & O'Hagan, A. (2001). Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society B*, **63**, 425–464. 19

Keys, A.C. & Rees, L.P. (2004). A sequential-design metamodeling strategy for simulation optimization. *Computers and Operations Research*, **31**, 1911–1932. 103

Khatri, C. & Shah, K.R. (1981). On the unbiased estimation of fixed effects in a mixed model for growth curves. *Communications in Statistics A-Theory and Methods*, **10**, 401–406. 68

Kilmer, R., Stinchcombe, M. & White, H. (1997). An emergency department simulation and a neural network metamodel. *Journal of the Society for Health System*, **5**, 63–79. 17

Kleijnen, J. (1987). *Statistical Tools for Simulation Practitioners*. Marcel Dekker, New York. 29

Kleijnen, J. (2008). *Design and Analysis of Simulation Experiments*. Springer, New York. 3, 19, 20, 31

Kleijnen, J., Beers, W.V. & Nieuwenhuyse, I.V. (2011). Expected improvement in efficient global optimization through bootstrapping. *Journal of Global Optimization*. 85

Kleijnen, J.P.C. (1998). *Experimental Design for Sensitivity Analysis, Optimization, and Validation of Simulation Models. Handbook of Simulation*, 173–223. Wiley, New York. 13, 29

Kleijnen, J.P.C. (2009). Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, **192**, 707–716. 61

Kleijnen, J.P.C. & Beers, W.C.M.V. (2005). Robustness of kriging when interpolating in random simulation with heterogeneous variances: Some experiments. *European Journal of Operational Research*, **30**, 102–119. 6, 7, 18, 19, 20, 21, 33, 45, 58, 59

Kleijnen, J.P.C. & Beers, W.C.M.V. (2010). Monotonicity-preserving bootstrapped kriging metamodels for expensive simulations. *Working paper*. 5

Kushner, H. & Clark, D. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer, New York, NY. 6, 84

Law, J. (1994). Introduction to neural networks: Design, theory, and applications. 17

Leoppky, J., Sacks, J. & Welch, W. (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, **51**, 366–376. 96

Li, R. & Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in gaussian kriging models. *Technometrics*, **47**, 111–120. 45

Li, Y.F., Ng, S.H., Xie, M. & Goh, T.N. (2010a). A systematic comparison of metamodeling techniques for simulation optimization in decision support systems. *Applied Soft Computing (to appear)*. 4, 6, 115

Li, Y.F., Ng, S.H., Xie, M. & Goh, T.N. (2010b). A systematic comparison of metamodeling techniques for simulation optimization in decision support systems. *Applied Soft Computing*, **10**, 1257–1273. 29

Liu, Y., Liew, K., Hon, Y. & Zhang, X. (2005). Numerical simulation and analysis of an electroactuated beam using a radial basis function. *Smart Materials and Structures*, **14**, 1163–1171. 16

Locatelli, M. (1997). Bayesian algorithm for one-dimensional global optimization. *Journal of the Global Optimization*, **10**, 57–76. 96, 100

Lophaven, S.N., Nielsen, H.B. & Songdergaard, J. (2002). Dace: a matlab kriging toolbox, version 2.0. *IMM Technical University of Denmark, Lyngby*. 42

Ma, C.X., Fang, K.T. & Lin, D.K.J. (2002). A note on uniformity and orthogonality. *Journal of Statistical Planning Inference*, **113**, 323–334. 24

Matheron, G. (1963). Principle of geostatistics. *Economic Geology*, **58**, 1246–1266. 4, 17, 19, 29

McDonald, D., Grantham, W., Tabor, W. & Murphy, M. (2007). Global and local optimization using radial basis functions response surface models. *Applied Mathematical Modeling*, **31**, 2095–2110. 16

McKay, M.D., Beckman, R.J. & Conover, W.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**, 239–245. 23

McMahon, J. & Franke, R. (1992). Knot selection for least squares thin plate splines. *SIAM Journal of Scientific and Statistical Computing*, **13**, 484–498. 15

Meckesheimer, M., Barton, R., Simpson, T., Limayemn, F. & Yannou, B. (2001). Metamodeling of combined discrete/continuous responses. *AIAA Journal*, **39**, 1950–1959. 16

Meckesheimer, M., Barton, R., Simpson, T. & Booker, A. (2002). Computationally inexpensive metamodel assessment strategies. *AIAA Journal*, **40**, 2053–2060. 16

Mitchell, T.J. & Morris, M.D. (1992). The spatial correlation function approach to response surface estimation. In J.J. Swain, D. Goldsman, R.C. Crain & J.R. Wilson, eds., *Proceedings of the 1992 Winter Simulation Conference*, 565–571, Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. 19

Miura, K. (2011). An introduction to maximum likelihood estimation and information geometry. *Interdisciplinary Information Sciences*, **17**, 155–174. 181

Mizutani, E. & Demmel, J. (2003). On structure-exploiting trust-region regularized nonlinear least squares algorithms for neural-network learning. *Neural Networks*, **16**, 745–753. 27

Mockus, J. (1994). Application of bayesian approach to numerical methods of global and stochastic optimization. *Journal of the Global Optimization*, **4**, 347–365. 5, 85

Mockus, J., Tiesis, V. & Zilinskas, A. (1978). *The application of Bayesian methods for seeking the extremum, Towards Global Optimisation, Vol. 2*, 117–129. North Holland, Amsterdam. 85

Morris, M. & Mitchell, T. (1995). Exploratory design for computational experiment. *Journal of Statistical Planning Inference*, **43**, 381–402. 25

Morris, M., Mitchell, T. & Ylvisaker, D. (1993). Bayesian design and analysis of computer experiments: use of derivative in surface prediction. *Technometrics*, **35**, 243–255. 18

Mosteller, F. & Tukey, J.W. (1968). Data analysis, including statistics. In *Handbook of Social Psychology*, Addison-Wesley, Reading, MA. 98

Munoz, J. & Felicisimo, A. (2004). Comparison of statistical methods commonly used in predictive modeling. *Journal of Vegetation Science*, **15**, 285–292. 15

Notteboom, T.E. & Vernimmen, B. (2009). The effect of high fuel costs on liner service configuration in container shipping. *Journal of Transport Geography*, **17**, 325–337. 106

Oakley, J. (2002). Eliciting gaussian process priors for complex computer codes. *The Statistician*, **51**, 81–97. 125

O'Hagan, A., Kennedy, M. & Oakley, J. (1999). Uncertainty analysis and other inference tools for comples computer codes (with discussion). In A.P.D. J. M. Bernardo J. O. Berger & A.F.M. Smith, eds., *Bayesian Statistics 6:Proceedings of the Sixth Valencia International Meeting*, 503–524, Oxford University Press, Oxford, UK. 19

Owen, A.B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, **2**, 439–452. 24

Patterson, H. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika*, **58**, 545–554. 72

Patterson, H. & Thompson, R. (1974). Maximum likelihood estimation of components of variance. In *In Proceedings of the 8th International Biometric Conference*, 197–207, Biometric Society, Washington, DC. 72

Pham, T. & Wagner, M. (1994). Filtering noisy images using kriging, signal processing and its applications. In J.D. Tew, S. Manivannan, D.A. Sadowski & A.F. Seila, eds., *Proceedings of the Fifth International Symposium*, 184–191, Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. 18

Picheny, V., Ginsbourger, D. & Richet, Y. (2010). Noisy expected improvement and on-line computation time allocation for the optimization of simulators with tunable precision. In *In Proceedings of the 2nd International Conference on Engineering Optimization*, 1–10, Lisbon, Portugal. 85, 94, 100, 104

Qian, P.Z.G. & Wu, C.F.J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, **50(2)**, 192–204. 124, 127, 128

Raiffa, H. & Schlaiffer, R. (2000). *Applied Statistical Decision Theory*. John Wiley, New York. 137

Rao, S. & Balakrishnan, N. (1999). Computational electromagnetics - a review. *Current Science*, **77**, 1343–1347. 2

Reinsel, G. (1984). Estimation and prediction in a multivariate random effects generalized linear model. *Journal of American Statistical Association*, **79**, 406–414. 68

Rocca, A.L. & Power, H. (2005). Free mesh radial basis function collocation approach for the numerical solution of system of multi-ion electrolytes. *International Journal for Numerical Methods in Engineering*, **64**, 699–734. 16

Ruppert, D. (2010). *Statistics and Data Analysis for Financial Engineering*. Springer, 1st edn. 13

Sacks, J., Welch, W.J., Mitchell, T.J. & Wynn, H.P. (1989). Design and analysis of computer experiments (with discussion). *Statistical Science*, **4**, 409–435. 4, 14, 17, 19, 24, 29, 139

Sakata, S., Ashida, F. & Zako, M. (2007). On applying kriging-based approximate optimization to inaccurate data. *Computer methods in applied mechanics and engineering*, **196**, 2055–2069. 30

Santner, T.J., Williams, B.J. & Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York. 3, 4, 18, 19, 23, 24, 121, 123, 124, 139, 184

Sasena, M., Papalambros, P. & Goovaerts, P. (2001). The use of surrogate modeling algorithms to exploit disparities in function computation time within simulation-based optimization. In *In Proceedings of the 4th Congress on Structural and Multidisciplinary Optimization*, ISSMO, Dalian, China. 19, 20, 28

Sasena, M., Papalambros, P. & Goovaerts, P. (2002). Exploration of metamodeling sampling criteria for constrained global optimization. *Engineering Optimization*, **34**, 263–278. 28

Shewry, M.C. & Wynn, H.P. (1987). Maximum entropy sampling. *Journal of Applied Statistics*, **14**, 165–170. 24

Shi, L. & Olafsson, S. (2000). Nested partitions method for global optimization. *Operations Research*, **48**, 390–407. 6

Simpson, T., Peplinski, J., Koch, P. & Allen, J. (2001). Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers*, **17**, 129–150. 4, 29

Stuart, A. & Andord, K. (1994). *Kendalls Advanced Theory of Statistics, Volume 1: Distribution Theory*. Wiley, New York. 179

Tekin, E. & Sabuncuoglu, I. (2004). Simulation optimization: A comprehensive review on theory and applications. *IIE Transactions*, **36**, 1067–1081. 6

Toyooka, Y. (1976). Prediction error in a linear model with estimated parameters. *Biometrika*, **69**, 453–459. 68

Wang, H., Li, E., Li, G.Y. & Zhong, Z.H. (2008). A metamodel optimization methodology based on multi-level fuzzy clustering space reduction strategy and its applications. *Computers and Industrial Engineering*, **55**, 503–532. 30

Wang, S., Chen, W. & Tsui, K.L. (2009). Bayesian validation of computer models. *Technometrics*, **51**, 439–451. 125

Wang, Y. & Fang, K.T. (1981). A note on uniform distribution and experimental design. *KeXue TongBao*, **26**, 485–489. 24

Watson, C. & Johnson, M. (2004). Hurricane loss estimation models: Opportunities for improving the state of the art. *Bulletin of the American Meteorological Society*, **85**, 1713–1726. 2

Welch, W. & Sacks, J. (1991). A system for quality improvement via computer experiments. *Communications in Statistics-Theory and Methods*, **20**, 477–495. 18

Welch, W., Yu, T., Kang, S. & Sacks, J. (1990). Computer experiments for quality control by parameter design. *Journal of Quality Technology*, **22**, 15–22. 17

Welch, W., Sacks, R., Sacks, J., Wynn, H., Mitchell, T. & Morris, M. (1992). Screening, predicting, and computer experiments. *Technometrics*, **34**, 15–25. 18

Woodbury, M.A. (1950). Inverting modified matrices. *Statistical Research Group, Princeton University, Princeton, N. J.*. 45

Wu, H. & Sun, F. (2007). Adaptive kriging control of discrete-time nonlinear systems. *Control Theory & Applications*, **1**, 646–656. 18

Yao, Z.S., Ng, S.H. & Lee, L.H. (2011). A study on shipping liner service plans with ship speed and bunkering port considerations. 2, 106, 107, 110

Yin, J., Ng, S.H. & Ng, K.M. (2008). Kriging model with modified nugget effect for random simulation with heterogeneous variances. In J.D. Tew, S. Manivannan, D.A. Sadowski & A.F. Seila, eds., *In Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management*, 1714–1718, Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. 6, 20, 21, 69, 75, 76, 81, 115, 119, 120, 122, 128

Yin, J., Ng, S.H. & Ng, K.M. (2009). A study on the effect of parameter estimation on the kriging model's prediction error in stochastic simulation. In M. Rossetti, R. Hill, B. Johnson, A. Dunkin & R. Ingalls, eds., *In Proceedings of the 2009 Winter Simulation Conference*, 674–685, Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey. 117, 121

Zimmerman, D. & Cressie, N. (1992a). Mean squared error prediction error in the spatial linear model with estimated cova-riance parameters. *Annals of the institute of statistical mathematics*, **44**, 27–43. 68

Zimmerman, D.L. & Cressie, N. (1992b). Mean squared error prediction error in the spatial linear model with estimated co-variance parameters. *Annals of the institute of statistical mathematics*, **44**, 27–43. 18, 81, 121

Appendix A

Kriging predictor and kriging variance for heteroscedastic model

According to Cressie (1993), the kriging predictor can be obtained by minimizing the mean squared error in Eq. (3.11). For the general case (using $Y(x_i)$ as the input to the model instead of $\bar{Y}(x_i)$), the MSE for the kriging predictor is

$$\begin{aligned}
MSE &= \mathrm{E}\left[P(Y(x_0)) - Y(x_0)\right]^2 \\
&= \mathrm{E}\left[\sum_{i=1}^{m} \lambda_i Y(x_i) - Y(x_0)\right]^2 \\
&= \mathrm{E}[(Y(x_0) - \mu)^2 + \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i\lambda_j (Y(x_i) - \mu)(Y(x_j) - \mu) \\
&- 2\sum_{i=1}^{m} \lambda_i (Y(x_0) - \mu)(Y(x_i) - \mu)] \\
&= var(Y(x_0)) + \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i\lambda_j cov(Y(x_i), Y(x_j)) - 2\sum_{i=1}^{m} \lambda_i cov(Y(x_i), Y(x_0))
\end{aligned}$$

given the kriging weight constraint $\sum_{i=1}^{m} \lambda_i = 1$, and from Eq. (3.15) $var(Y(x_0)) = cov(Y(x_0), Y(x_0)) = \iota_0^* + \iota_1$; $cov(Y(x_i), Y(x_j)) = \iota_1 corr(d_{ij})$. Then minimizing

$$\mathrm{E}\left[\sum_{i=1}^{m} \lambda_i Y(x_i) - Y(x_0)\right]^2 - 2l\left(\sum_{i=1}^{m} \lambda_i - 1\right)$$

with respect to $\lambda_i : i = 1, 2, \ldots, m$ and the Lagrange multiplier l gives the kriging predictor as follows:

$$P(Z(x_0)) = \sum_{i=1}^{m} \lambda_i \bar{Y}(x_i)$$

with

$$\lambda_i = c^{\mathrm{T}} R^{-1} e_i + F^{\mathrm{T}} R^{-1} \frac{(1 - F^{\mathrm{T}} R^{-1} c)^{\mathrm{T}}}{F^{\mathrm{T}} R^{-1} F} e_i$$

where R is given in Eq. (3.16). The kriging variance is given as follows:

$$MSE(x_0) = \iota_{x_0}^* + \iota_1 \left[1 - c^{\mathrm{T}} R^{-1} c \frac{\left(1 - F^{\mathrm{T}} R^{-1} c\right)^2}{F^{\mathrm{T}} R^{-1} F}\right]$$

Considering the decomposition of $R = R_Z + R_\varepsilon$, we can have the result in Equation (3.18).

Appendix B

MSE for the modified nugget-effect and nugget-effect model

The MSE_S at the ith observation point for the kriging model with modified nugget-effect is

$$
\begin{aligned}
MSE_S(x_i) &= \iota_1 \left[1 - (c + F\frac{(1 - F^{\mathrm{T}}R'^{-1}c)}{F^{\mathrm{T}}R'^{-1}F})^{\mathrm{T}}R'^{-1}c + \frac{(1 - F^{\mathrm{T}}R'^{-1}c)}{F^{\mathrm{T}}R'^{-1}F}\right] - \iota^*_{x_i} \\
&= \iota_1[1 - (c + F\frac{(1 - F^{\mathrm{T}}R^{-1}c) + F^{\mathrm{T}}R^{-1}(R^{-1} + \eta^{-1})^{-1}R^{-1}c}{F^{\mathrm{T}}R'^{-1}F})^{\mathrm{T}}R'^{-1}c \\
&+ \frac{(1 - F^{\mathrm{T}}R^{-1}c) + F^{\mathrm{T}}R^{-1}(R^{-1} + \eta^{-1})^{-1}R^{-1}c}{F^{\mathrm{T}}R'^{-1}F}] - \iota^*_{x_i} \\
&= \iota_1[1 - (c + F\frac{F^{\mathrm{T}}R^{-1}(R^{-1} + \eta^{-1})^{-1}R^{-1}c}{F^{\mathrm{T}}R'^{-1}F})^{\mathrm{T}}R'^{-1}c \\
&+ \frac{F^{\mathrm{T}}R^{-1}(R^{-1} + \eta^{-1})^{-1}R^{-1}c}{F^{\mathrm{T}}R'^{-1}F}] - \iota^*_{x_i} \\
&= \iota_1[1 - (c + F\frac{F^{\mathrm{T}}R^{-1}(R^{-1} + \eta^{-1})^{-1}\eta^{-1}\eta R^{-1}c}{F^{\mathrm{T}}R'^{-1}F})^{\mathrm{T}}R'^{-1}c \\
&+ \frac{F^{\mathrm{T}}R^{-1}(R^{-1} + \eta^{-1})^{-1}\eta^{-1}\eta R^{-1}c}{F^{\mathrm{T}}R'^{-1}F}] - \iota^*_{x_i} \\
&= \iota_1[1 - (c + F\frac{F^{\mathrm{T}}R'^{-1}\eta e_i}{F^{\mathrm{T}}R'^{-1}F})^{\mathrm{T}}R'^{-1}c + \frac{F^{\mathrm{T}}R'^{-1}\eta e_i}{F^{\mathrm{T}}R'^{-1}F}] - \iota^*_{x_i}
\end{aligned}
$$

Let $\Delta_m = F^{\mathrm{T}} R'^{-1} F$ indicate the summation of all the elements in the inverse correlation matrix, and let Δ_{mi} represent the summation of the ith column or row. Then we have

$$\begin{aligned}
MSE_S(x_i) &= \iota_1 \left[1 - (c + F\frac{\eta_i \Delta_{mi}}{\Delta_m})^{\mathrm{T}} R'^{-1} c + \frac{\eta_i \Delta_{mi}}{\Delta_m} \right] - \iota^*_{x_i} \\
&= \iota_1 [1 - 1 - \sum_{j=1}^{m} c_j \eta_i (\frac{\Delta_{mi}\Delta_{mj}}{\Delta_m} - \Delta_m) + \frac{\eta_i \Delta_{mi}}{\Delta_m}] - \iota^*_{x_1} \\
&= \iota_1 [-\eta_i (\frac{\Delta_{mi}}{\Delta_m} - 1) \sum_{j=1}^{m} c_j \delta_{mj} + \frac{\eta_i \Delta_m i}{\Delta_m}] - \iota^*_{x_1} \\
&= \iota_1 [-\eta_i (\frac{\Delta_{mi}}{\Delta_m} - 1)(1 - \eta_i \Delta_{mi}) + \frac{\eta_i \Delta_m i}{\Delta_m}] - \iota^*_{x_1} \\
&= \iota_1 [-\eta_i (\frac{\Delta_{mi}}{\Delta_m} - 1 - \eta_i \frac{\Delta^2_{mi}}{\Delta_m} + \eta_i \Delta_i) + \frac{\eta_i \Delta_m i}{\Delta_m}] - \iota^*_{x_1} \\
&= \iota_1 [\eta_i + \eta_i^2 \frac{\Delta^2_{mi}}{\Delta_m} - \eta_i^2 \Delta_{mi}] - \iota^*_{x_1}
\end{aligned}$$

For the modified nugget-effect model, $\eta_i = \iota^*_{x_i}/\iota_1$, so

$$MSE_S(x_i) = \iota_1 (\eta_i^2 \frac{\Delta^2_{mi}}{\Delta_m} - \eta_i^2 \Delta_{mi})$$

Appendix C

Proof for the two-stage algorithm

Assuming a single dimension optimization problem can be given as finding

$$Z^*(x) = \min_{x \in [x_a, x_b]} Z(x)$$

Here x is the input variable and $Z(x)$ is the response. x_a and x_b are the lower and upper bound of the design space. Based on the algorithm given in Section 5.4.3, we consider the case when the total computing budget T is large enough and the stopping criterion for the algorithm is $mEI(x^*) < l$, and x^* indicates the current best location that maximizes the mEI function.

Now assuming we have n observations $Z(x_1), Z(x_2), ..., Z(x_n)$ at locations $x_1, x_2, ..., x_n$, for the mEI function values at these observed locations, we have

$$\begin{aligned} mEI(x_i)_{i=1,2,...,n} &= (Z_{min}(x) - \hat{Z}_m(x_i))\Phi\left(\frac{Z_{min}(x) - \hat{Z}_m(x_i)}{s_Z(x_i)}\right) \\ &+ s_Z(x_i)\phi\left(\frac{Z_{min}(x) - \hat{Z}_m(x_i)}{s_Z(x_i)}\right) \end{aligned}$$

Here $\hat{Z}_m(x_0)$ indicates the MNEK predictor, and $s_Z(x_0)$ is the standard deviation of spatial uncertainty. Since the spatial uncertainty at the observed location is zero, so $s_Z(x_i)_{i=1,2,...,n} = 0$. It can easily be shown that $s_Z(x_i)\phi\left(\frac{Z_{min}(x) - \hat{Z}_m(x_i)}{s_Z(x_i)}\right) = 0$ $i = 1, 2, ..., n$. Since $Z_{min}(x)$ is the current best value, we have $Z_{min}(x) <$

$\hat{Z}_m(x_i)_{i=1,2,...,n}$, which suggests that $\Phi\left(\frac{Z_{min}(x)-\hat{Z}_m(x_i)}{s_Z(x_i)}\right)=0$ and hence

$$mEI(x_i)=0 \quad i=1,2,...,n$$

Now we know that the mEI function equals to zero on the design points $x_1, x_2, ..., x_n$. According to the algorithm in Section 5.4.3, the stopping criterion of the algorithm is $mEI(x^*) \leq l$, which suggests that the algorithm will not sample in the region with mEI function value lower than l.

Given that

$$\frac{\partial mEI(x)}{\partial s_Z(x)}=\phi\left(\frac{Z_{min}(x)-\hat{Z}_m(x_0)}{s_Z(x)}\right)>0$$

mEI function is monotonic increasing in $s_Z(x)$. Moreover, we know that the spatial uncertainty $s_Z^2(x_0)$ equals to zero when $x_0=x_i,\ i=1,2,...,n$, and increases as the minimum distance between the prediction point x_0 and existing design point x_i, $d_{0i}=|x_0-x_i|_{i=1,2,...,n}$ increases. Hence we can identify a small region D_i around the existing design point x_i where $\forall x_0 \in D_i$, $mEI(x_0)<l$. The length of this small region can be given as

$$\tilde{d}_i=|x_{ui}-x_{li}|$$

where x_{ui} and x_{li} satisfy

$$\frac{s_Z(x_{li})}{\sqrt{2\pi}}=\frac{s_Z(x_{ui})}{\sqrt{2\pi}}=\frac{1}{\sqrt{2\pi}}\sigma_Z^2\left[1-c(x_{ui})^{\mathrm{T}}R_Z^{-1}c(x_{ui})\frac{(1-F^{\mathrm{T}}R_Z^{-1}c(x_{ui}))^2}{F^{\mathrm{T}}R_Z^{-1}F}\right]=l$$

. Since

$$\begin{aligned}
mEI(x_0)_{i=1,2,...,n} &= (Z_{min}(x)-\hat{Z}_m(x_0))\Phi\left(\frac{Z_{min}(x)-\hat{Z}_m(x_0)}{s_Z(x_0)}\right)\\
&+ s_Z(x_0)\phi\left(\frac{Z_{min}(x)-\hat{Z}_m(x_0)}{s_Z(x_0)}\right)\\
&\leq s_Z(x_0)\phi\left(\frac{Z_{min}(x)-\hat{Z}_m(x_0)}{s_Z(x_0)}\right)\\
&\leq s_Z(x_0)\phi(0)=\frac{s_Z(x_0)}{\sqrt{2\pi}}
\end{aligned}$$

We can derive that for all the x_0 in D_i bounded by x_{li} and x_{ui}, $mEI(x_0)\leq l$.

So given stopping criterion l, the maximum number of design points will be a finite value of

$$n_{max} = \frac{b-a}{\min\limits_{i=1,2,\ldots,n} \tilde{d}_i(l)}$$

When $l \to 0$, $x_{li} = x_{ui} \to x_i$, and hence $\tilde{d}_i \to 0$, $n_{max} \to \infty$ Given the mEI function can visit any possible unobserved location in $[a, b]$, we have

$$\lim_{n\to\infty} \max_{i=1,2,\ldots,n} (x_i - x_{i-1}) = 0$$

The observed set of design points is dense in $[a, b]$, we can show that

$$\lim_{n\to\infty} Z_n^* = Z^*$$

According to the algorithm, once the $mEI(x^*) \leq l$, the algorithm will stop observing new design points and distribute all the remaining computing budget to the existing design points. Since we adopt the OCBA for the Allocation Stage, this can guarantee the algorithm will converge to the optimum by the end when $n \to \infty$. The convergence property of OCBA can be found in Chen *et al.* (2000).

Appendix D

Details of the two-point example

For the two point example in Figure 6.3, with the linear correlation function and the constant underlying mean function for the Z process, we have

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_Z^2 & \sigma_Z^2\rho_0 \\ \sigma_Z^2\rho_0 & \sigma_Z^2 \end{pmatrix} + \begin{pmatrix} \sigma_Z^2\frac{\tau_1}{r_0} & 0 \\ 0 & \sigma_Z^2\frac{\tau_2}{r_0} \end{pmatrix},$$

$$v^{\mathrm{T}} = \sigma_Z^2(\frac{1+\rho_0}{2}, \frac{1+\rho_0}{2}),$$

$$F^{\mathrm{T}} = (1,1), F(x_0) = 1.$$

Suppose an equal number of replications r_0 are taken at both design points and $D_n^{\mathrm{T}} = [Y_n^{\mathrm{T}}, S_n^{2\mathrm{T}}]$, where

$$Y_n^{\mathrm{T}} = (y_1, y_2),\ S_n^{2\mathrm{T}} = (s_1^2, s_2^2).$$

With homogeneous variance and sample variances approximately $s_1^2 = s_2^2 = s_0^2$,

$$S_n^{2\mathrm{T}} = (s_0^2, s_0^2).$$

Assuming the following non informative priors (Jeffreys Prior)

$$p(\beta) = 1, p(\sigma_Z^2) = 1/\sigma_Z^2, p(\tau) = \frac{1}{2(1+\tau)^2}$$

the posterior predictive distribution of $Z(x_0)$ can be obtained following Equation (6.6).

$$\begin{aligned} f(Z(x_0)|D_n) &= \iiint f(Z(x_0)|\beta,\tau,\sigma_Z^2,D_n)f(\beta|\sigma_Z^2,D_n)f(\tau|\sigma_Z^2,D_n)f(\sigma_Z^2|D_n) \\ &\quad d\beta d\sigma_Z^2 d\tau \\ &\propto \iiint f(Z(x_0)|\beta,\tau,\sigma_Z^2,D_n)f(Y_n|\beta,\tau,\sigma_Z^2)f(s_0^2|\beta,\tau,\sigma_Z^2)f(\beta|\sigma_Z^2) \\ &\quad f(\tau|\sigma_Z^2)f(\sigma_Z^2)d\beta d\sigma_Z^2 d\tau \end{aligned}$$

where the likelihoods $f(Y_n|\beta,\tau,\sigma_Z^2)$ follows a normal distribution and $f(s_0^2|\beta,\tau,\sigma_Z^2)$ follows the Pearson type III distribution, see Chapter 11, Stuart & Andord (1994).

$$f(s_0^2) \propto \frac{r_0}{2\tau\sigma_Z^2}^{(r_0-1)/2}(s_0^2)^{(r_0-3)/2}\exp\left(\frac{-r_0 s_0^2}{2\tau\sigma_Z^2}\right)$$

With $f(\beta|\sigma_Z^2) \propto 1$, $f(\tau|\sigma_Z^2) \propto \frac{1}{2(1+\tau)^2}$,

$$\begin{aligned} f(Z(x_0)|D_n) &\propto \iiint (\sigma_Z^2)^{-1/2}\exp(-\frac{1}{2}\frac{(Z(x_0)-F(x_0)\beta-c^{\mathrm{T}}(x_0)R^{-1}(Y_n-F\beta))^2}{\sigma_Z^2(1-c^{\mathrm{T}}(x_0)R^{-1}c(x_0))})\cdot \\ &\quad (\sigma_Z^2)^{-1}\exp(-\frac{1}{2}(Y_n-F\beta)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(Y_n-F\beta))\cdot \\ &\quad (\sigma_Z^2)^{-\frac{r_0+1}{2}}\exp(-\frac{r_0 s_0^2}{2\tau\sigma_Z^2})\frac{1}{2(1+\tau)^2}d\beta d\sigma_Z^2 d\tau \end{aligned}$$

Following the simplification steps in Appendix F, where

$$\begin{aligned} p &= Z(x_0)-c^{\mathrm{T}}(x_0)R^{-1}Y_n = \frac{-r_0(y_1+y_2-2Z(x_0))(1+\rho_0)+2Z\tau}{2(r_0+r_0\rho_0+\tau)} \\ k &= c^{\mathrm{T}}(x_0)R^{-1}F-F^{\mathrm{T}}(x_0) = -\frac{\tau}{r_0+r_0\rho_0+\tau} \\ \mathbb{A} &= \left(\frac{k^{\mathrm{T}}k}{1-c^{\mathrm{T}}(x_0)R^{-1}c(x_0)}+F^{\mathrm{T}}R^{-1}F\right) = \frac{2(r_0-r_0\rho_0+\tau)}{r_0-r_0\rho_0^2+2\tau} \\ \mathbb{B} &= \left(\frac{p^{\mathrm{T}}k}{1-c^{\mathrm{T}}(x_0)R^{-1}c(x_0)}+Y_n^{\mathrm{T}}R^{-1}F\right) = \frac{r_0(y_1+y_2)(1-\rho)+2Z\tau}{r_0(1-\rho_0^2)+2\tau} \end{aligned}$$

we get the posterior predictive distribution as

$$\begin{aligned} f(Z(x_0)|D_n) &\propto \int \sqrt{2\pi\mathbb{A}^{-1}}(\frac{1}{2\sigma_Z^2}\mathbb{B}^{\mathrm{T}}\mathbb{A}^{-1}\mathbb{B}-\frac{1}{2}(\frac{p^{\mathrm{T}}p}{\sigma_Z^2-v^{\mathrm{T}}\Sigma^{-1}v}+Y_n^{\mathrm{T}}\Sigma^{-1}Y_n))^{-1/2} \\ &\quad d\sigma_Z^2 d\tau \\ &\propto \left(1+\frac{(Z(x_0)-\frac{y_1+y_2}{2})^2}{2(\frac{(y_1-y_2)^2}{2}+\frac{s_0^2}{r_0}+\frac{s_0^2 r_0(1-\rho_0)}{2[\ln(r_0(1-\rho_0))(1-\rho_0)r_0+r_0\rho_0]})}\right)^{-\frac{1+r_0}{2}} \end{aligned}$$

Appendix E

Estimating Predictor Variance by Delta Method

Chapter 6 provides a Bayesian perspective of the kriging metamodeling for stochastic simulation in order to better account for the parameter estimation uncertainties. Alternative approaches can be used to estimate the predictor's uncertainties with unknown parameters, such as Delta method.

As stated in Section 6.2.1, the typical unknown parameters for the kriging metamodeling can be given as $\theta = (\beta, \sigma_Z^2, \phi_Z, \sigma_\xi^2)$. If we use MLE to estimate all these parameters, the estimators $\hat{\beta}, \hat{\sigma}_Z^2, \hat{\phi}_Z, \hat{\sigma}_\xi^2$ will be asymptotically normal. The delta method can be used to derived an approximation of the probability distribution of the predictor considering it as a function of all these estimated parameters, see Agresti (2002) for the details on delta method.

Based on the predictor form in Eq. (3.17), we can write the predictor as:

$$\hat{Z}(x_0, \hat{\beta}, \hat{\sigma}_Z^2, \hat{\phi}_Z, \hat{\sigma}_\xi^2) = F(x_0)\hat{\beta} + \hat{\sigma}_Z^2 c(x_0, \hat{\phi}_Z)^{\mathrm{T}} (\hat{\sigma}_Z^2 R_Z(\hat{\phi}_Z) + \hat{\sigma}_\xi^2 R_\varepsilon)^{-1} (\bar{Y} - F\hat{\beta}) \tag{E.1}$$

Given the MLE of these estimator are asymptotically normal and

$$\hat{\theta} = (\hat{\beta}, \hat{\sigma}_Z^2, \hat{\phi}_Z, \hat{\sigma}_\xi^2),$$

we could have the approximation of the predictor variance shown in

$$\mathrm{Var}(\hat{Z}(x_0,\hat{\theta})) = \nabla\hat{Z}(\hat{\theta})^{\mathrm{T}}\frac{\Sigma_\theta}{n}\nabla\hat{Z}(\hat{\theta}) \tag{E.2}$$

where $\hat{\theta} \sim N(\theta^*, \Sigma_\theta)$, θ^* is the true value of the unknown parameter and Σ_θ is the covariance function.

For the simple two point problem in Section 6.2.3.2, we assume that the parameters ϕ_Z are known. Moreover, we know that the estimators for the unknown parameters have asymptotically normal distribution $\hat{\theta} \sim N(\theta^*, \Sigma_\theta)$, and Σ_θ can be given as the inverse of Fisher information matrix on θ when some of the regularity conditions are satisfied, see Miura (2011) for details. Based on the assumptions of the unknown parameters β, σ_Z^2 and σ_ξ^2, we have the Fisher information matrix as

$$I(\theta) = I\begin{pmatrix}\beta\\ \sigma_Z^2\\ \sigma_\xi^2\end{pmatrix} = \begin{pmatrix} -E\left(\frac{\partial^2 l}{\partial\beta\partial\beta}\right) & -E\left(\frac{\partial^2 l}{\partial\beta\partial\sigma_Z^2}\right) & 0\\ -E\left(\frac{\partial^2 l}{\partial\sigma_Z^2\partial\beta}\right) & -E\left(\frac{\partial^2 l}{\partial\sigma_Z^2\partial\sigma_Z^2}\right) & 0\\ 0 & 0 & -E\left(\frac{\partial^2 l}{\partial\sigma_\xi^2\partial\sigma_\xi^2}\right)\end{pmatrix}$$

The predictor can be given as

$$\hat{Z}(x_0) = \hat{\beta} + \hat{\sigma}_Z^2\begin{pmatrix}\frac{1+\rho_0}{2} & \frac{1+\rho_0}{2}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\hat{\sigma}_Z^2+\hat{\sigma}_\xi^2/r_0 & \hat{\sigma}_Z^2\rho_0\\ \hat{\sigma}_Z^2\rho_0 & \hat{\sigma}_Z^2+\hat{\sigma}_\xi^2/r_0\end{pmatrix}\begin{pmatrix}y_1-\hat{\beta}\\ y_2-\hat{\beta}\end{pmatrix}$$

The predictor variance can be derived as

$$\begin{aligned}\mathrm{Var}(\hat{Z}(x_0)) &= \nabla\hat{Z}(\hat{\theta},x_0)^{\mathrm{T}}\frac{I^{-1}}{2}\nabla\hat{Z}(\hat{\theta},x_0)\\ &= \frac{(1-\rho_0)\Delta_y}{2r_0(1+\rho_0)} + \frac{s_0^2}{r_0}\\ &+ \frac{\sqrt{r_0^2(1-\rho_0^2)(r_0\Delta_y-4s_0^2)-16s_0^2(1+\rho_0)(4s_0^2+r_0\Delta_y(1-\rho_0))}}{2r_0(1+\rho_0)}\end{aligned}$$

where $\Delta_y = (y_1-y_2)^2$. Compared with the results in Equation (6.7), we can see that the predictor variance given by Delta method still consists of three components, representing the uncertainties from observations y_1 and y_2, random noise s_0^2 and the mixed effect of estimating σ_Z^2 and σ_ξ^2. However the mixed effect component is slightly different from the results in Section 6.2.3.2, this is probably due to the prior setting used in Section 6.2.3.2.

Appendix F

Proof for Proposition 1

If τ is known, the posterior distribution in (6.6) can be given as

$$
\begin{aligned}
f(Z(x_0)|Y_n) &= \iint f(Z(x_0)|Y_n,\beta,\sigma_Z^2) f(Y_n|\beta,\sigma_Z^2) f(\beta|\sigma_Z^2) f(\sigma_Z^2) d\beta d\sigma_Z^2 \\
&\propto \iint (\sigma_Z^2)^{-1/2} \exp(-\frac{1}{2}\frac{(Z(x_0) - F^{\mathrm{T}}(x_0)\beta - v^{\mathrm{T}}(x_0)\boldsymbol{\Sigma}^{-1}(Y_n - F\beta))^2}{\sigma_Z^2 - v^{\mathrm{T}}(x_0)\boldsymbol{\Sigma}^{-1}v(x_0)}) \\
&\quad (\sigma_Z^2)^{-(p+1)/2} \exp(-\frac{1}{2}(Y_n - F^{\mathrm{T}}\beta)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(Y_n - F^{\mathrm{T}}\beta)) \\
&\quad (\sigma_Z^2)^{-\alpha_z - n/2 - 1} \exp(-\frac{1}{2}(\beta - w_0)^{\mathrm{T}}\frac{Q_0^{-1}}{\sigma_Z^2}(\beta - w_0)) \exp(-\frac{\gamma_z}{\sigma_Z^2}) d\beta d\sigma_Z^2 \\
&= \iint \exp(-\frac{1}{2\sigma_Z^2}\beta^{\mathrm{T}}\mathbb{A}\beta + \frac{1}{\sigma_Z^2}\mathbb{B}^{\mathrm{T}}\beta)(\sigma_Z^2)^{-\alpha_z - n/2 - 3/2 - (p+1)/2} \\
&\quad \exp(-\frac{1}{2}(\frac{{}^{\prime 2}}{\sigma_Z^2 - v^{\mathrm{T}}(x_0)\boldsymbol{\Sigma}^{-1}v(x_0)} + Y_n^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}Y_n + w_0^{\mathrm{T}}\frac{Q_0^{-1}}{\sigma_Z^2}w_0 + \frac{2\gamma_z}{\sigma_Z^2})) \\
&\quad d\beta d\sigma_Z^2 \\
&= \int (\sigma_Z^2)^{-\alpha_z - n/2 - 3/2} \sqrt{\det 2\pi\mathbb{A}^{-1}} \exp(\frac{1}{2\sigma_Z^2}\mathbb{B}^{\mathrm{T}}\mathbb{A}^{-1}\mathbb{B} \\
&- \frac{1}{2}(\frac{\mathbf{p}^2}{\sigma_Z^2 - v^{\mathrm{T}}(x_0)\boldsymbol{\Sigma}^{-1}v(x_0)} \\
&+ Y_n^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}Y_n + w_0^{\mathrm{T}}\frac{Q_0^{-1}}{\sigma_Z^2}w_0 + \frac{2\gamma_z}{\sigma_Z^2})) d\sigma_Z^2
\end{aligned}
$$

where

$$\begin{aligned}
\mathbb{A} &= (\frac{\mathbf{k}^{\mathrm{T}}\mathbf{k}}{1-c^{\mathrm{T}}(x_0)R^{-1}c(x_0)} + F^{\mathrm{T}}(x)R^{-1}F(x) + Q_0^{-1}) \\
\mathbb{B} &= (\frac{\mathbf{p}^{\mathrm{T}}\mathbf{k}}{1-c^{\mathrm{T}}(x_0)R^{-1}c(x_0)} + Y_n^{\mathrm{T}}R^{-1}F + w_0Q_0^{-1}) \\
\mathbf{p} &= Z(x_0) - c^{\mathrm{T}}(x_0)R^{-1}Y_n, \mathbf{k} = c^{\mathrm{T}}(x_0)R^{-1}F - F^{\mathrm{T}}(x_0) \\
\mathbf{\Sigma} &= \sigma_Z^2 R = \sigma_Z^2(R_Z + R_\xi) \\
v(x_0) &= \sigma_Z^2 c(x_0)
\end{aligned}$$

Collecting all terms involving σ_Z^2, we have

$$\begin{aligned}
f(Z(x_0)|Y_n) &\propto \int (\sigma_Z^2)^{-\alpha_z - n/2 - 3/2} \exp(-\frac{\Omega}{\sigma_Z^2}) d\sigma_Z^2 \\
&\propto \Omega^{-\alpha_z - n/2 - 1/2} \\
&\propto \left(1 + \frac{1}{2\alpha_z + n} \frac{(Z(x_0) - \mu_p(x_0))^2}{\sigma_p^2(x_0)}\right)^{-\alpha_z - n/2 - 1/2}
\end{aligned}$$

where

$$\begin{aligned}
\Omega &= (\frac{1}{2}\mathbb{B}^{\mathrm{T}}\mathbb{A}^{-1}\mathbb{B} - \frac{1}{2}(\frac{\mathbf{p}^2}{1-c^{\mathrm{T}}(x_0)R^{-1}c(x_0)} + Y_n^{\mathrm{T}}R^{-1}Y_n + w_0^{\mathrm{T}}Q_0^{-1}w_0 + 2\gamma_z) \\
\mu_p(x_0) &= F(x_0)\lambda/M + c^{\mathrm{T}}(x_0)R^{-1}(Y_n - F\lambda/M) \\
\sigma_p^2(x_0) &= \frac{2\gamma_z + Y_n^{\mathrm{T}}R^{-1}Y_n + w_0^{\mathrm{T}}Q_0^{-1}w_0 - \lambda^{\mathrm{T}}\lambda/M}{2\alpha_z + n}(1 - c^{\mathrm{T}}(x_0)R^{-1}c(x_0) + \mathbf{k}^{\mathrm{T}}\mathbf{k}/M) \\
\lambda &= F^{\mathrm{T}}R^{-1}Y_n + w_0Q_0^{-1}, M = F^{\mathrm{T}}R^{-1}F + Q_0^{-1}
\end{aligned}$$

Appendix G

Posterior distribution of the parameters

To derive the distribution $f(\sigma_\xi^2(x_0)|\mathbf{S}_n^2, \phi_V)$, we first consider the distribution of the log of the noise level. As we place an independent Gaussian process prior over the log of the noise level, if assuming the observations $\mathbf{S}_n^2$ are sufficiently accurate to ignore the sampling noise and an exact interpolation model is adequate, then from the results in Santner *et al.* (2003), p. 95, the conditional distribution $f(V_{x_0}|\mathbf{S}_n^2, \phi_V)$ is a non central t distribution with mean μ_V and covariance $\mathbf{\Sigma}_V$. Assuming V_n follows multivariate normal distribution with mean $F\beta_V$ and covariance matrix $\sigma_V^2 R_V$, and let $c_V(x_0)$ and R_V represent the correlation vector and the correlation matrix (which are functions of ϕ_V), and further assume prior distributions for parameters β_V and σ_V^2:

$$f(\beta_V, \sigma_V^2, \phi_V) = f(\beta_V|\sigma_V^2) f(\sigma_V^2) f(\phi_V)$$

$$\begin{aligned} f(\beta_V|\sigma_V^2) &\sim N_p(w_V, \sigma_V^2 Q_V) \\ f(\sigma_V^2) &\sim IG(\alpha_V, \gamma_V) \\ f(\phi_V) &\sim \exp(-b_V\phi_V)(\phi_V)^{a_V-1} \end{aligned}$$

Then the conditional can be given as

$$f(V_{x_0}|\mathbf{S}_n^2, \phi_V) \propto T_1(n + \alpha_V, \mu_V, \mathbf{\Sigma}_V)$$

where

$$\mu_V = F(x_0)^{\mathrm{T}} \frac{F^{\mathrm{T}} R_V^{-1} \mathbf{S}_n^2 + Q_V^{-1} b_V}{F^{\mathrm{T}} R_V^{-1} F + Q_V^{-1}} + c_V(x_0) R_V^{-1} (\mathbf{S}_n^2 - \frac{F^{\mathrm{T}} R_V^{-1} \mathbf{S}_n^2 + Q_V^{-1} b_V}{F^{\mathrm{T}} R_V^{-1} F + Q_V^{-1}})$$

$$\begin{aligned} \mathbf{\Sigma}_V &= \frac{2\gamma_V + \kappa^{\mathrm{T}}((F^{\mathrm{T}} R_V^{-1} F)^{-1} + Q_0)^{-1} \kappa}{n + 2\alpha_V} \\ &+ \frac{\mathbf{S}_n^2 (R_V^{-1} - R_V^{-1} F (F^{\mathrm{T}} R_V^{-1} F)^{-1} F^{\mathrm{T}} R_V^{-1}) \mathbf{S}_n^2}{n + 2\alpha_V} \\ &\quad (1 - c_V^{\mathrm{T}} R_V^{-1} c_V + \frac{(c_V^{\mathrm{T}} R_V^{-1} F - F(x_0))^{\mathrm{T}} (c_V^{\mathrm{T}} R_V^{-1} F - F(x_0))}{F^{\mathrm{T}} R_V^{-1} F}) \end{aligned}$$

$$\kappa = b_V - \frac{F^{\mathrm{T}} R_V^{-1} \mathbf{S}_n^2}{F^{\mathrm{T}} R_V^{-1} F}$$

Since $V_{x_0} = \ln(\sigma_\xi^2(x_0))$, with $f(V_{x_0}|\mathbf{S}_n^2, \phi_V)$, the conditional distribution can be obtained as

$$f(\sigma_\xi^2(x_0)|\mathbf{S}_n^2, \phi_V) = \frac{f(V_{x_0}|\mathbf{S}_n^2, \phi_V)}{\sigma_\xi^2(x_0)}$$

The conditional distribution of the parameters ϕ_V can be given as

$$\begin{aligned} f(\sigma_V^2|\mathbf{S}_n^2, \phi_V) &\propto \int f(\mathbf{S}_n^2|\beta, \sigma_V^2, \phi_V) f(\beta_V|\sigma_V^2) f(\sigma_V^2) d\beta_V \\ f(\phi_V|\mathbf{S}_n^2, \sigma_V^2) &\propto \int f(\mathbf{S}_n^2|\beta, \sigma_V^2, \phi_V) f(\beta_V|\sigma_V^2) f(\phi_V) d\beta_V \end{aligned}$$

Given the prior distribution in (G.1), the conditional distribution of the pa-

rameters σ_V^2 and ϕ_V can be given as

$$\begin{aligned}
f(\sigma_V^2|\mathbf{S}_n^2,(\phi_V)^i) &\propto (\det(\sigma_V^2 M_V((\phi_V)^i)R_V((\phi_V)^i)))^{-1/2}\exp(\frac{1}{2\sigma_V^2}\lambda_V((\phi_V)^i)^{\mathrm{T}} \\
&\quad M_V^{-1}((\phi_V)^i) \\
&\quad \lambda_V((\phi_V)^i) - \frac{1}{\sigma_V^2}S_n^{2\mathrm{T}}R_V^{-1}((\phi_V)^i)\mathbf{S}_n^2 - w_V^{\mathrm{T}}\frac{Q_V^{-1}}{\sigma_V^2}w_V \\
&- \frac{\gamma_V}{\sigma_V^2})(\sigma_V^2)^{1/2-\alpha_V}\exp(-b_V(\phi_V)^i) \\
f(\phi_V|\mathbf{S}_n^2,(\sigma_V^2)^{i-1}) &\propto (\det((\sigma_V^2)^{i-1}M_V R_V))^{-1/2}\exp(\frac{1}{2(\sigma_V^2)^{i-1}}\lambda_V^{\mathrm{T}}M_V^{-1}\lambda_V \\
&- \frac{1}{(\sigma_V^2)^{i-1}}S_n^{2\mathrm{T}}R_V^{-1}\mathbf{S}_n^2 \\
&\quad -w_V^{\mathrm{T}}\frac{Q_V^{-1}}{(\sigma_V^2)^{i-1}}w_V - \frac{\gamma_V}{(\sigma_V^2)^{i-1}})\exp(-b_V\phi_V)(\phi_V)^{a_V-1}
\end{aligned}$$

where

$$\lambda_V(\phi_V) = F^{\mathrm{T}}R_V^{-1}(\phi_V)\mathbf{S}_n^2 + w_V Q_V^{-1},\ M_V(\phi_V) = F^{\mathrm{T}}R_V^{-1}(\phi_V)F + Q_V^{-1}$$

The conditional distribution of the parameters σ_Z^2 and ϕ_Z can be given as

$$\begin{aligned}
f(\sigma_Z^2|Y_n,\phi_Z) &\propto \int f(Y_n|\beta,\sigma_Z^2,\phi_Z)f(\beta|\sigma_Z^2)f(\sigma_Z^2)d\beta \\
f(\phi_Z|Y_n,\sigma_Z^2) &\propto \int f(Y_n|\beta,\sigma_Z^2,\phi_Z)f(\beta|\sigma_Z^2)f(\phi_Z)d\beta
\end{aligned}$$

With the prior distributions in (6.8), the integrand above can be simplified to

$$\begin{aligned}
f(Y_n|\beta,\sigma_Z^2,\phi_Z) &\propto (\det\boldsymbol{\Sigma})^{-1/2}\exp(-\frac{1}{2}(Y_n - F\beta)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(Y_n - F\beta)) \\
f(\beta|\sigma_Z^2) &\propto (\sigma_Z^2)^{-1/2}\exp(-\frac{1}{2}(\beta - w_0)^{\mathrm{T}}\frac{Q_0^{-1}}{\sigma_Z^2}(\beta - w_0)) \\
f(\sigma_Z^2) &\propto \exp(-\frac{\gamma_z}{\sigma_Z^2})(\sigma_Z^2)^{-\alpha_z+1} \\
f(\phi_Z) &\propto \exp(-b_z\phi_Z)(\phi_Z)^{a_z-1}
\end{aligned}$$

Therefore the conditionals can be obtained as

$$\begin{aligned} f(\sigma_Z^2|Y_n,(\phi_Z)^i) &\propto (\det(\sigma_Z^2 M((\phi_Z)^i)R((\phi_Z)^i)))^{-1/2}\exp(\frac{1}{2\sigma_Z^2}\lambda((\phi_Z)^i)^{\mathrm{T}} \\ &\quad M^{-1}((\phi_Z)^i)\lambda((\phi_Z)^i) - \frac{1}{\sigma_Z^2}Y_n^{\mathrm{T}}R^{-1}((\phi_Z)^i)Y_n \\ &\quad -w_0^{\mathrm{T}}\frac{Q_0^{-1}}{\sigma_Z^2}w_0 - \frac{\gamma_z}{\sigma_Z^2})(\sigma_Z^2)^{1/2-\alpha_z}\exp(-b_z(\phi_Z)^i) \\ f(\phi_Z|Y_n,(\sigma_Z^2)^{i-1}) &\propto (\det((\sigma_Z^2)^{i-1}MR))^{-1/2}\exp(\frac{1}{2(\sigma_Z^2)^{i-1}}\lambda^{\mathrm{T}}M^{-1}\lambda \\ &- \frac{1}{(\sigma_Z^2)^{i-1}}Y_n^{\mathrm{T}}R^{-1}Y_n \\ &\quad -w_0^{\mathrm{T}}\frac{Q_0^{-1}}{(\sigma_Z^2)^{i-1}}w_0 - \frac{\gamma_z}{(\sigma_Z^2)^{i-1}})\exp(-b_z\phi_Z)(\phi_Z)^{a_z-1} \end{aligned}$$

where

$$\lambda(\phi_Z) = F^{\mathrm{T}}R^{-1}(\phi_Z)Y_n + w_0Q_0^{-1},\ M(\phi_Z) = F^{\mathrm{T}}R^{-1}(\phi_Z)F + Q_0^{-1}$$

As for the direct sampling from these posterior distributions can be difficult, we adopt the rejection-acceptance method for the posterior distribution sampling within the Gibbs loop for the MCMC implement.

Appendix H

Posterior distribution of σ_Z^2

The posterior distribution of σ_Z^2 can be given as follows

$$\begin{aligned} f(\sigma_Z^2|y_D) &= \int f(y_D|\beta,\sigma_Z^2)f(\beta|\sigma_Z^2)f(\sigma_Z^2)d\beta \\ &= \int \det(\mathbf{\Sigma})^{-1/2}\exp\left[-\frac{1}{2}(y_D-F\beta)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(y_D-F\beta)\right] \\ &\quad \exp\left[-\frac{1}{2}(\beta-w_0)^{\mathrm{T}}\frac{Q_0^{-1}}{\sigma_Z^2}(\beta-w_0)-\frac{\gamma_z}{2\alpha_z}\right](\sigma_Z^2)^{1/2-\alpha_z-(1+p)/2}d\beta \\ &\propto \det(M\mathbf{\Sigma})^{-1/2}\exp[\frac{1}{2\sigma_Z^2}\lambda^{\mathrm{T}}M^{-1}\lambda-\frac{1}{2\sigma_Z^2}y_D^{\mathrm{T}}R^{-1}y_D-w_0^{\mathrm{T}}\frac{Q_0^{-1}}{\sigma_Z^2}w_0 \\ &- \frac{\gamma_z}{\sigma_Z^2}](\sigma_Z^2)^{1/2-\alpha_z} \\ &= (\frac{A_G}{\sigma_Z^2})^{-\alpha_z-n/2-3/2}\exp\left[-\frac{B_G}{\sigma_Z^2}\right] \end{aligned}$$

where

$$A_G = \det((F^{\mathrm{T}}R_DF)R_D)^{-1/2}$$

$$B_G = \frac{2\gamma_z + w_0^{\mathrm{T}}Q_0^{-1}w_0 + y_D^{\mathrm{T}}R_D^{-1}y_D - \lambda_D^{\mathrm{T}}M_D^{-1}\lambda_D}{2}$$

Therefore, the posterior distribution is an inverse gamma distribution with mean $\left[\frac{2\gamma_z+y_D^{\mathrm{T}}R_D^{-1}y_D+w_0^{\mathrm{T}}Q_0^{-1}w_0-\lambda_D^{\mathrm{T}}M_D^{-1}\lambda_D}{2\alpha_z+n}\right]$, where

$$\lambda_D = F_D^{\mathrm{T}}R_D^{-1}y_D + w_0Q_0^{-1},\ M_D = F_D^{\mathrm{T}}R_D^{-1}F_D + Q_0^{-1}.$$